U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 1931

Evaluation of a Game-Based Simulation During Distributed Exercises

Michael J. Singer and Bruce W. Knerr
U. S. Army Research Institute

September 2010

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS
Director

Technical review by:

John S. Barnett, U.S. Army Research Institute
Rodney Long, U.S. Army Research, Development, and Engineering Command – Simulation and Training Technology Center

NOTICES

DISTRIBUTION: Primary distribution of this Research Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 2511 Jefferson Davis highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Research Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE: September 2010	2. REPORT TYPE: Final	3. DATES COVERED: January 2008 – March 2009

4. TITLE AND SUBTITLE: Evaluation of a Game-Based Simulation During Distributed Exercises	5a. CONTRACTOR OR GRANT NUMBER: 5b. PROGRAM ELEMENT NUMBER: 622785
6. AUTHOR(S): Michael J. Singer & Bruce W. Knerr U.S. Army Research Institute 12350 Research Parkway Orlando, FL, 32826	5c. PROJECT NUMBER: A790 5d. TASK NUMBER: 294 5e. WORK UNIT NUMBER: H01
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-IF 12350 Research Parkway, Orlando, FL, 32826	8. PERFORMING ORGANIZATION REPORT NUMBER:
9. SPONSORING / MONITORING ABENCY NAME(S) AND ADDRESS(ES): U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, VE 22202-3926	10. MONITOR ACRONYM: ARI 11. MONITOR REPORT NUMBER: Research Report 1931

12. DISTRIBUTION / AVAILABILITY STATEMENT:

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES:

Subject Matter POCs: Michael J. Singer and Bruce W. Knerr

14. ABSTRACT *(Maximum 200 words)*:

Two exercises using a Game-Based Simulation (GBS) were conducted by the U.S. Army Research Development and Engineering Command, Simulation and Training Technology Center (RDECOM-STTC) and the United Kingdom Land Warfare Development Group. Soldiers from the U.S. Army and the U.K. military conducted coalition mission rehearsals during each exercise. Data were collected on the system user interface, on the effectiveness of unit and joint exercise sessions, and on After Action Review (AAR) functionality and applications. Several issues in technological capabilities limited and constrained the military tasks that could be performed during the exercises, and limited the AARs. Nevertheless, questionnaire data collected during each exercise indicated several positive aspects of using game-based simulations. The GBS system was considered capable of providing considerable scope for general dismounted Soldier rehearsal and training. The graphics and user interface were judged adequate for use in training rehearsals and AARs, especially in preparation for home station field training exercises. The largest negative issue was the limited number of weapon types and equipment. The second largest issue was the limited equipment functionality that the system supported. A third issue was the lack of sufficient numbers of civilians and opposing forces for different interactions in the non-kinetic exercises.

15. SUBJECT TERMS:

Game-Based Simulation, Training, Simulation, User Evaluation, Distributed Rehearsal

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	75	Ellen Kinzer Technical Publication Specialist (703) 602-8049

Research Report 1931

Evaluation of a Game-Based Simulation During Distributed Exercises

Michael J. Singer
Bruce W. Knerr
U. S. Army Research Institute

Training Technology Research Unit
Stephen Goldberg, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

September 2010

Army Project Number	**Personnel, Performance**
622785A790	**and Training Technology**

Approved for public release; distribution is unlimited.

ACKNOWLEDGMENT

This data collection exercise could not have been conducted without the collaboration and support of RDECOM-STTC and The Technical Cooperation Panel. Also key in the conduct of the exercises as well as the development and collection of data were Amanda Suprenant, Grant Taylor, and Dr. Christian Jerome.

EVALUATION OF A GAME-BASED SIMULATION DURING DISTRIBUTED EXERCISES

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army is making a substantial commitment to the use of Game-Based Simulations (GBSs) for training, readiness, and concept development, as well as test and evaluation. Although these systems are used for a wide range of operations, opportunities for evaluation are limited. In addition, these systems can enable larger operations with greater numbers of Soldiers, and should enable distributed coalition forces to rehearse together. The On-Line Interactive Virtual Environment system (OLIVE, Forterra Systems, Inc.) was modified under contract to the Research, Development, and Engineering Command, Simulation and Training Technology Center (RDECOM-STTC) to investigate the feasibility and effectiveness of providing realistic training and rehearsal for large groups of dismounted Soldiers conducting a wide range of primarily non-kinetic operations. The U.S. Army Research Institute for the Behavioral & Social Sciences (ARI) addressed a major research challenge within the project by working to identify and quantify the effects of game-based system capabilities, characteristics, and features on learning, skill acquisition, retention, and transfer for U.S. Army tasks.

Procedure:

Exercises were structured between the United Kingdom Land Warfare Development Group and RDECOM-STTC in order to evaluate technology for a distributed multiplayer GBS, and the U.S. Army Research Institute for the Behavioral & Social Sciences (ARI) was asked to gather information on the potential training effectiveness. These exercises were designed as coalition mission rehearsals for platoon (minus) groups connected via the internet. Simulation laboratories were established at RDECOM-STTC in Orlando, FL and at the Defence Academy of the United Kingdom, Shrivenham, GB to support the exercises. In the first event, Cadets from West Point and Officers from Ft. Benning participated from the Orlando laboratory, and Soldiers from the 3rd Mercians (U.K.) participating from Shrivenham conducted coalition missions over four days. Several months later, Soldiers from the 10th Mountain Division (U.S. Army) and a different group of Soldiers from the 3rd Mercians (U.K.) conducted another set of coalition missions for four days. During both exercise events, data were collected on the system user interface after initial training on system use. Exercise questionnaires addressing system characteristics and training potential were administered following some of the instructional and exercise sessions. Questionnaires addressing the After Action Review (AAR) functionality and application were administered following the final exercise AAR at each event. Additional questionnaires and measures were also administered to collect information addressing all participant's current game-play experience and self-rated expertise.

Findings:

Both exercise events were structured to investigate and demonstrate the technology capabilities rather than address specific coalition training goals. Several different technical

issues with the OLIVE prototype limited and constrained the military tasks that could be performed during the planned exercise missions. Nevertheless, questionnaire data collected during each exercise event indicated several positive and negative aspects of using the GBS. The graphics and user interface systems were judged as adequate for use in rehearsals, despite the limited equipment functionality (primarily weapons and vehicles). The OLIVE prototype was also judged as providing considerable scope for general dismounted Soldier rehearsal and training. The questionnaire responses also indicated that Soldiers found the system easier to work with than the more logistically difficult real-world (live) training and rehearsal activities. In addition, the best and most functional aspect of the system was the ability to provide AAR supporting replays and static visuals. The biggest negative issue was the lack of supporting equipment that Soldiers use during training and mission accomplishment. Without the complete range of weapons, communication equipment, and vehicles, it was difficult for Soldiers to address even the non-kinetic aspects of general military operations. In addition, the lack of "clutter" (e.g., civilians and opposing forces) in the environment during operations seemed to emphasize those missing informational aspects rather than the possibilities for interaction that did exist.

In spite of systemic deficiencies, the information gathered during the two episodes demonstrated that exercises can be conducted with widely dispersed contingents, and information on effects can be acquired. This type of GBS is usable by military personnel engaged in military activities (even if non-doctrinal). Further, the Soldiers involved accepted the GBS and perceived some benefit from the exercises. Soldiers also seemed to accept that this type of GBS can be used for training at their home stations.

Utilization and Dissemination of Findings:

The U. S. Army will employ game-based simulation technology for training, mission planning, rehearsal, and constructive test and evaluation. RDECOM-STTC is continuing to address the evolving technological capabilities of game-based simulation systems. Based upon the engineering information gathered, in conjunction with the Soldier evaluations of the system capabilities, RDECOM-STTC and ARI are continuing to collaborate in the development and evaluation of game-based simulations. The usability and effectiveness results from these initial efforts are being used to shape further GBS development, employment, and evaluation efforts.

Understanding the user interface, functionality, AAR functionality and training effectiveness will contribute to effective specification of GBS training configurations for different uses. The information ARI has generated has been used to plan a third coalition exercise, as well as structure long-range plans for integrating dismounted Soldier ground simulations with integrated coalition air support simulations.

CONTENTS

CONTENTS (CONTINUED)

Page

LIST OF TABLES

EVALUATION OF A GAME-BASED SIMULATION DURING DISTRIBUTED EXERCISES

Currently, the U.S. Army trains Soldiers to perform conventional warfare tasks through schoolhouse courses, unit-based training, and live training events at Combat Training Centers (CTC). Institutional courses (e.g., Basic Officer Leaders Course) take considerable time and effort to alter in response to changes in the Contemporary Operational Environment (COE). The schoolhouse can teach doctrine for various Soldier roles, but cannot address the wide range of Standard Operating Procedures (SOPs) that are developed at the unit level. Mission and sustainment training that incorporates SOPs and unit Tactics, Techniques, and Procedures (TTPs) is conducted at the unit's home location, which often has limited ranges, Military Operations in Urban Terrain (MOUT) sites, and training support. Current lessons learned from the COE are incorporated into mission exercises conducted at the CTC, which work very hard to maintain currency. In short, establishing and supporting training for the increasingly non-kinetic (no or limited weapons), geographically and culturally-centered missions, SOPs, and unit determined TTPs has become a central focus for the U. S. Army.

The U.S. Army is currently fielding a Game-Based Simulation (GBS) that allows Soldiers to train TTPs, perform mission planning and rehearsal operations, and practice decision-making tasks against current enemy tactics (Bohemia Interactive. 2009). The fielded system is Virtual BattleSpace 2: Army (VBS2; Bohemia Interactive. 2009). While virtual training for Dismounted Infantry (DI) has lagged behind that of vehicles because of the complexity of the multiple team tasks and the levels of interaction between the avatars (graphical representations of users interacting physically in a virtual world), VBS2: Army will allow training and rehearsal of kinetic missions with a limited number of DI trainees (apparently less than 50) at each site (Robson, 2008).

Game-Based Simulation Development and Evaluation

Prior to the fielding of VBS2: Army, a different system was being developed and investigated by the U.S. Army's Research, Development and Engineering Command, Simulation and Training Technology Center (RDECOM-STTC) with the support and collaboration of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). That research effort has leveraged and adapted a commercial massively multi-player online game (OLIVE™, which stands for “Online Interactive Virtual Environment, produced by Forterra Systems, Inc.) as a simulation for dismounted infantry (DI) training and rehearsal (Singer, et al., 2008). The focus of the development effort was to provide an easy-to-use, internet-based simulation that leaders can use to train and rehearse new TTPs for responding to asymmetric threats, especially situations not based on kinetic operations. Although the emphasis in this GBS was on non-kinetic aspects of dismounted Soldier operations, equipment and weapons that support military operations were also incorporated to a limited extent (Singer, et al., 2008). The goal, starting in 2003, was to develop a simulation that could become a training multiplier when used in conjunction with field exercises. The intent has always been to provide a powerful tool that augments and supports Situational Training Exercises (STX), without replacing "boots on the ground" training.

ARI has collaborated with RDECOM-STTC in this program (Singer, et al., 2008) as part of a larger ARI program, GamBIT (Assessing and Improving the Effectiveness of Game-Based Simulations). The goal of GamBIT is to improve the training effectiveness of game-based simulations through the development of needed training capabilities and investigation of the applicability of the simulations to collective training objectives. ARI's role in the METER program was to investigate the usability, acceptability, and potential training effectiveness of the GBS in order to provide information for future technology development as well as for Army acquisition and fielding programs. ARI activities have included conducting formative evaluations that focused on training effectiveness, needed fidelity, and instructional tools for training in the COE. During the engineering development and adaptation phases, ARI conducted formative evaluations of the system through administering questionnaires and conducting interviews with participating Soldiers during demonstration exercises. The primary use of information accrued during that effort was to adapt or alter the system engineering or interface design in order to support increased usefulness by potential users. The post-exercise information indicated that the Soldiers and leaders considered the system capable of preparing them for more expensive and time-consuming live exercises (Singer, et al., 2008).

The OLIVE system was initially selected because there were very few systems that were constructed as Massively Multiplayer Online Game (MMOG) engines. Some background concepts are presented next as framing information or concepts. In order to be clear about the system being evaluated, a minimal exposition of "massively multiplayer," "persistent," and "non-kinetic" is in order.

"Massively multiplayer" refers to an online game system with a large number of concurrent users interacting in the virtual world. Many systems, such as "World of Warcraft", support more than a thousand users in the virtual world at one time over the internet. Most games in this genre use a client-server architecture, with each user machine referred to as a "client" on the system. The servers set the world for the clients, and update all clients based on the changes made by each individual client. This provides a distinction with systems that are exclusively "peer to peer," meaning that each user's machine sends information to all the other users on a local area network. In order to control the load on the server and network, the area represented and the number of clients that access each area are limited. In OLIVE, the areas are limited, but the number of users that are able to interact within each area is supposed to be very large (on the order of hundreds). The number of users and detail of the environment does require the use of relatively high end machines.

"Persistent" means that the environment represented on the server and updated on each client persists over time. If an object is left in the environment, it remains in the environment (barring server or "world" resets) until removed. A weapon or vehicle that is "dropped" will stay there until picked up, put away, or driven off. This does not mean that bullet holes or explosions will remain, as they are limited duration graphics objects.

The goal of the original program was to investigate game capabilities that are needed beyond the weaponry that the military already has simulated in other venues (e.g., the Engagement Skills Trainer 2000). "Non-kinetic" is a label that encompasses this concept, as the

goal of a non-kinetic simulation is not to train the use of any particular weapon, but to address the decisions and interactions between Soldiers, and with civilians.

The concept of non-kinetic exercises took advantage of the emerging capabilities of the OLIVE system. The system was designed to support larger numbers of participants, enabling more realistic urban terrains for non-kinetic operations rehearsals. In addition, OLIVE enabled individuation - easily developed differences in the appearance of the user "avatar" or human animations (gestures) within the system. The non-kinetic concept also pushed the incorporation of improved voice interactions within the simulation using Voice Over Internet Protocols (VoIP), which in OLIVE provides both localized speech and several radio systems, each with multiple selectable channels. Initial engineering and operational evaluations during the development of the OLIVE system seemed to support all of these needed capabilities. Input from early evaluations also led to the development of an After Action Review (AAR) capability. The OLIVE AAR system was essentially a record and replay approach with video-like controls. During later development, a maneuverable viewpoint was adapted, with functionality that allowed distributed trainees to join a trainer controlled replay. This enabled a controller to show specific segments of a recorded scenario from a single point of view to a large number of distributed personnel.

For RDECOM-STTC, the technological issues leading to the joint exercises were whether the system could be used with relatively large numbers of operators and large amounts of equipment in widely distributed exercises. The engineering and computational capabilities were key factors of interest in conducting their efforts. For ARI, the focus was on precursor skills and knowledge enabling Soldiers to participate, and the acquisition of Soldier input on the usability, practicality, and effect on training during non-kinetic operations as the central goals for rehearsal. These exercises, hosted and supported by RDECOM-STTC, provided an opportunity to gather subjective information and evaluations on the potential training effectiveness of large scale, distributed exercises for individual Soldier rehearsal and training.

Coalition Mission Exercises

It is likely that current operations and future deployments will shift more towards multi-national coalitions that will conduct coordinated missions with limited and short-term goals. Future conflicts will likely arise quickly, and may require small-unit joint operations with little time for training or rehearsal, according to the Office of the Undersecretary of Defense for Acquisition, Technology and Logistics report on training for future conflicts (OUSDATL, 2003). These operations will have to coordinate and cooperate with local culture, politics, institutions, and resources (OUSDATL, 2003). This future scenario may present problems for organizations that have typically worked with separately defined goals and areas. Should coalition partners be called upon to integrate forces in any operation, training would be needed at the basic unit level in accordance with the Army principle of "Train as you Fight" (FM7-0, 2002). In addition, the United Kingdom is establishing a laboratory at the Land Warfare Centre (LWC) in Warminster, U.K., for investigating and testing virtual exercises in a joint task force framework up to U.K. Brigade and U.S. Division levels. As a result, the LWC were also interested in the system requirements for distributed, non-kinetic operations rehearsal and the potential effects on training and rehearsal.

These exercises were initiated as a cooperative research effort under The Technical Cooperation Program, Training Technology Technical Panel 2 (TTCP, TTTP2; http://www.dtic.mil/ttcp/). As noted on the website, the TTCP organization was formed under agreements between Australia, Canada, New Zealand, the United Kingdom, and the United States to cooperate and participate in defense scientific and technical information exchange and collaborative research projects. The goal of the U.S./U.K. conducted Coalition Mission Experimental Exercises (CMEX) was to evaluate the use of the OLIVE technology for distributed training and mission rehearsals. As a part of the U.K.'s investigation into using gaming technology for training at the LWC, and based on their membership in the TTCP, the LWC agreed to begin with small unit coalition exercises.

The RDECOM-STTC program supporting these exercises is the Multinational Experimentation for Training, Evaluation and Research (METER). The intent of the METER program is to address many of the issues referred to above; 1) investigating the engineering requirements for distributed small unit-based operations, 2) investigating the training equipment and training methods for small unit coalition forces working cooperatively, and 3) providing experience with a U.S. Army developmental program for a cooperating nation while continuing the evaluation work on the RDECOM-STTC developmental software. The program goal is to run a series of exercises, starting with ground exercises and working toward Close Air Support (CAS) exercises in conjunction with existing multi-national joint aircraft simulation networks. Negotiations and planning for the exercises reported here was initiated in the fall of 2007, with agreement established in January 2008. The first test mission was scheduled for July of 2008 (referred to as Coalition Mission Exercise One, CMEX-I), and was used as a rehearsal for conducting a test with a coherent unit of U.S. Soldiers in October 2008 (referred to as CMEX-II). The next experiment is planned for the fall of 2009.

The overall approach combines engineering and software tests with a scenario-based training session that could be used to evaluate the potential for training effectiveness. This approach was made more challenging by the ongoing development of the GBS prototype for the distributed exercises. The staged exercises were designed to support a crawl-walk-run approach that would provide reliable and comprehensive data that could be used to provide information supporting specifications for the development and fielding of future GBS's.

Questionnaires

The major research issues of interest were: the usability of the GBS interface, potential training effectiveness, and support for feedback. In addition, some standard biographical information needed to be collected, information on computer expertise and game experience was needed, and a check on all participants' state of health was administered. Based on prior experience with the GBS being used (OLIVE, Singer, et al., 2008), we decided that existing questionnaires could be used to collect information in these areas with minimal adaptation. As the required information for the two exercises was the same, the initial plan was to keep the measures essentially the same in each exercise.

During the interval between the exercises, minor alterations were made to the Graphical User Interface, the Exercise Questionnaire (examining fidelity and training effect potential), and the AAR questionnaire (addressing the functions and effectiveness of the AAR) in order to move to a different data collection software system.. During that process, minor edits were also made in an attempt to improve clarity and decrease redundancy. The differences are shown in the appendices which describe the questions and illustrate the response categories or characteristics. Each of these questionnaires contained items that were intended to form coherent scales addressing important aspects of the system being evaluated. Those scales were constructed through consultation with subject matter experts in game development, Soldiers experienced in using simulations for training, and the management team guiding development of the software. The questionnaires have not been used enough to collect data sufficient for factor analyses. (For example, an online tutorial by Hinkin, 1998 recommends that questionnaires should have an item or question to observation ratio of at least one to four before conducting a factor analysis or reliability analysis.)

Game performance assessment battery. One measure for any technology-based training simulation effort is the time required to achieve proficiency in using the simulation. A common lesson learned from previous evaluations of games as training media has been that insufficient time was allowed for trainees to learn to operate and become proficient with the game-based simulation prior to using the system for training. Another tendency has been to over-estimate prior gaming experience and proficiency possessed by the participants. In addition, a major issue in establishing the effect of prior game experience or skills is that these issues were simply addressed with self-reports and self-ratings rather than any kind of objective measures. These issues were the central focus of the development of the Game Performance Assessment Battery (GamePAB; Chertoff, Jerome, Martin, & Knerr, 2008; Taylor, Singer, & Jerome, 2009).

The developers decided to use simulated tasks and related knowledge questions instantiated in the GamePAB (Chertoff, Jerome, Martin, & Knerr, 2008; Taylor, Singer, & Jerome, 2009) to quantify the gaming experience and skill of the exercise participants (trainees, role players, and controllers). While the system is still in development, it was used in these exercises to investigate possible differences between participating groups. The use of GamePAB during these exercises is a first step toward gathering the data necessary to establish the measure as reliable and valid. The intent is to enable investigations of possible relationships with objective exercise performance measures or the acquisition of GBS skills required before training is initiated.

GamePAB requires users to perform common tasks in a game framework, and collects data about the performance of those tasks. One segment of GamePAB requires participants to move their avatar through the environment while manipulating posture and movement speed, and communicating verbally (answering questions about the environment). A second segment in the game environment requires tracking a moving target, and hitting that target based upon a color cue. The output provides several response time measures (e.g., Posture Reaction Time and Communication Reaction Time) as well as accuracy data (e.g., Percentage of Correct Communications and Percent Time On Track). The Posture Reaction Time measures the time required to mimic an automated guide, changing posture as the guide does while traversing a route. The Communication Reaction Time measures the time for correct responses to questions

about the environment, answered while traversing the guided route. The Percentage of Correct Communications reflects the number of questions answered correctly, and the Percent Time on Track reflects the amount of time following within a criterion distance of the automated guide during route movement. There are also target tracking and firing accuracy measures (Percent Aim Time and Shot Reaction Time).

Game experience measure. The Game Experience Measure (GEM; Chertoff, et al., 2008; Appendix B) was developed to investigate participant's self-ratings of experiences with a wide range of games, and then to actually test their knowledge of games. The intent with this measure is to investigate the ratings of experience separately from an actual test of knowledge about specific game situations and controls from popularly rated games. The theme in development was to attempt to separate the experience and skill that people claim in general from the actual correct knowledge about games that would be expected from highly experienced game players.

The Game Experience scale addresses general gaming habits, frequency of play in different genres, experience with the user's favorite games, and user experience with different game controllers. General questions address the respondent's general gaming habits (scoring more highly if confidence and playing time were high in general). Other questions address the frequency of game play with specific genres; with more play overall contributing to a higher score. Several questions address expertise with the respondents' favorite games, and a number of questions address expertise with different types of game controllers. The scale is calculated by averaging the overall 5-point Likert scale responses for all the questions.

The Game Knowledge scale is assessed through a series of questions addressing six relatively recent and popular video games. Participants view a screenshot from each game and then answer multiple choice questions addressing controls required for specific actions and likely non-player responses based on the situation portrayed in the screenshot. The scale is then generated as the percent of correct responses.

One aspect of the ARI research program is to determine the relationship (if any) between our initial GamePAB measures of proficiency (e.g., movement skill, weapon aim/tracking skill, communication skill), with the GEM experience and knowledge outcomes. These measures can then be investigated for relationships with potential train-up requirements for GBS exercises. The initial effort will be an examination of any differences between the participating groups, evaluation of the measures, and investigation of any relationships with the other questionnaires. A long term goal is to establish objective performance measures within a GBS and investigate potential relationships with the two measures. Only the initial effort can be addressed in this report, in conjunction with possible relationships with the conceptually based exercise questionnaire.

Graphical user interface questionnaire. One major ongoing issue in the development of new systems based on computer use is the usability of the controlling interface. A poor interface, one difficult to manipulate in simulating task performance, will both inhibit user acceptance and diminish the potential for learning and transfer. The Graphical User Interface (GUI) questionnaire (see Appendix C) was derived from user questionnaires developed during

the initial evaluations of the OLIVE system (Singer, et al., 2008). Four scales were developed from the prior questions, as well as newly developed questions designed to address a wider range of issues. The conceptually developed scales (comprised of overlapping sets of questions) address the fidelity of the user interface, avatar capabilities, training issues, and general control operations.

The GUI Fidelity scale specifically addressed the realism of buildings and interactions with them (e.g., entering and searching), avatar appearance and movement capabilities (including identification and rank), as well as communications and gestures. Other included questions addressed searching or using menus, and specifically addressed system latency and realism. As shown in Appendix C, the response scales focused on difficulty of use and realism in use. Although the avatar capability questions were all included in the Fidelity scale, they were also interesting in isolation. The GUI Avatar Scale was derived from those questions addressing avatar capabilities and controls, moving and representing gestures, as well as the avatar's recognizability.

Another group of questions addressed the controls and operations within the GBS. These addressed ease of understanding and quality of the user interface design, including the ease of use in function shortcuts and capabilities as well as controls for movement, view manipulation, search functions and voice/radio communications. The last group of questions addressed support for training or goal accomplishment. These few questions asked about or required a direct response to training capabilities inherent in the user interface. They addressed whether menus could work for training purposes, whether the avatar appearance would support training, and whether the system supported military authority (e.g., chain of command) in achieving mission goals.

Exercise questionnaire. One primary measure of system capabilities and potential use can be derived from the trainee/participant responses to questions addressing the training effect and fidelity of the system relevant to the mission(s) performed. This is, admittedly, not as reliable nor as valid as direct performance measures of changed behaviors should be. However, presuming some accurate self-knowledge on the part of participants who have had far more than minimal training in military exercises, the subjective evaluations should provide some information about the system. An exercise questionnaire minimally modified from previous application with the same system (Singer, et al., 2008) was used to address the capabilities, functions, and issues relevant to the effective use of the system in transition training (for example, between mission training/rehearsal and full mission rehearsal in the field exercises).

The Exercise Questionnaire's general format uses a question or question stem that addresses some aspect of the GBS and provides a five or seven point Likert response scale with anchors (see Appendix D for question stems, and response scales). The questions generally fall into two categories: Fidelity issues and Training Effectiveness issues. As noted above, the items and scales were developed through evaluation by knowledgeable GBS users and developers, with input from Soldiers familiar with the simulation. The questionnaire responses are combined to form the two scales presented in the results and discussions.

The Fidelity scale questions addressed the GBS functions and capabilities, in terms of the task requirements in the exercises. Questions addressed the range of avatar capabilities including: gesturing, movement, visual and physical inspection, and equipment use. Other fidelity questions addressed the quality of sounds in general, any noticeable system latency, the quality of special effects (e.g., explosions), the adequacy of terrain/environment representation, movement realism, the quality of voice communications (both local and radio aspects), and equipment (e.g., vehicle use, binoculars, weapons).

The Training Effectiveness scale from the Exercise Questionnaire addressed comparisons with field training and preparation, estimates of training effectiveness, and evaluations of equipment functionality that were presumed to affect training. These questions required ratings comparing the GBS supported exercises to exercises conducted using maps, terrain mockups, or actual field exercises. Some of the questions directly addressed exercise preparation compared to other rehearsal or training situations (e.g., map rehearsals). This scale included some of the questions used in the fidelity scale, as the adequacy of representation also contributes to training by supporting detection, selection, and operations during military tasks. Following this logic, the Training Effectiveness scale also included ratings of the adequacy of avatar capabilities and representation of sounds in the environment. Questions in the Training Effectiveness scale also addressed perceived skill changes in the individual respondent and evaluations of team performance.

AAR questionnaire. A key aspect of the effectiveness of training using the OLIVE system (or any GBS for training) is AAR effectiveness. The most rigorous approach, of course, would be to compare the performance of units on similar (or identical) scenarios before and after AARs covering the same material with the different presentation methods. Since that was outside the scope of these efforts, we decided to address the AAR activities using more formative techniques. The easiest method for investigating the effects of the AAR system was to question the Soldiers involved on areas considered to be the most relevant to training effectiveness resulting from the administration of an AAR: the interface and general training feedback presentation capabilities.

The AAR questionnaire items conceptually clustered into scales addressing the AAR Interface and the AAR Training Capability. The AAR Interface scale addresses ease of understanding and use, the avatar capabilities in the AAR, sounds and voice, and ease of presentation. The AAR Training Capability scale used some of the same questions, in combination with other issues, to derive an evaluation of whether the AAR system could be used to support training. The questions addressed comparisons with field training presentations, preparation time and effort, as well as AAR capabilities and determining areas for improvement.

Biographical questionnaire and ancillary information. In prior related efforts, we have typically obtained background biographical information about the prior experience and training of participants (Singer, et al., 2008). Much of this baseline information was collected through an adapted biographical and computer experience questionnaire. Versions of this questionnaire have been used with prior ARI research (Fober, et al., 2001). The information gathered is simple and general non-personally identifying information that can be used to reference the participants to the overall military or civilian population, e.g., age, education, time

in career, and experience with computer programs in general. The computer use information gathered is distinctly different than the game referenced questionnaires, as it references common work programs like Microsoft Excel[tm] or Word[tm], or software programming languages like C++ or Java.

One of the issues with technology-based simulation is the accumulation of deleterious side-effects (Stanney & Kennedy, 1997; Kennedy, et al., 1992). While these effects have been found to some extent with virtual environment systems that completely replace the normal visual display, they have not been documented for PC/game console use. Nevertheless, as a part of the institutional review process, it was suggested that some measure of discomfort be incorporated into the approach. Therefore, a previously developed Simulator Sickness questionnaire (SSQ; Kennedy, et al., 1992) was used to obtain ratings of any side effects that arise while Soldiers perform tasks in the training event.

Finally, it is difficult to encompass all aspects of each factor that may affect user acceptance, system capabilities, or training effectiveness with questionnaires. As an attempt to capture areas or factors that have been missed in the questionnaires we recorded interviews with probing questions after the exercises were completed. The interview goal was to gain some insights that might be missed in the questionnaires, and support discovery of critical factors that may have been overlooked by the non-military data collectors.

Coalition Mission Exercise One

The primary goals of the first exercise were to verify the logistical capability to conduct distributed exercises between the United Kingdom and the United States and to establish a baseline of information regarding the operation of the system and its potential for training. The technology should be capable of internationally distributed exercises without excessive time lags or technical problems, given that commercial games seem to manage while using different time-frames and long distance communications. The coordination of distributed military training exercises does require a steep learning curve, and the initial effort was intended to generate needed lessons learned for conducting further experiments/training exercises. Primarily, the initial data collection provided an opportunity to test out the questionnaires and measures of performance, as well as providing information about differences in U.S./U.K. AAR techniques. The planned sequence for the exercises is presented in Table 1. The table glosses over the timing difficulties caused by the five hour time difference between the U.S. and U.K. Lead time was also required for the coordination of the leaders and trainers conducting the exercises, scenario development, equipment preparation, and support staff (e.g., role-players) training.

From the outset the approach taken in CMEX I was constrained by the perceived needs of the military personnel recruited to participate in the exercises, and the engineering requirements for a large-scale network that were established by RDECOM-STTC. In addition, several system constraints combined to limit the number of participants and the possible roles that those participants could exercise within the system. As noted above, the OLIVE system was developed to focus on the non-kinetic aspects of dismounted Soldier operations, and therefore had a limited amount of weaponry available for use. In addition, the number of participants at each location were targeted for platoon-minus (an incomplete platoon, without any heavy

weapons). The available network connections, as well as number and capability of computers, also contributed to the limitations on the number of participants at each location. In addition to the participants, the network had to support exercise controllers, semi-automated forces (SAF) computer systems, and role-players. The role-players were needed to provide situational stimuli for decision-making by the Soldiers. Finally, the servers that were used to support the exercises were located in California, complicating the network connections and control.

Table 1. Proposed Timeline for Exercises

	Activity	Data Collection
1st Day	Introduction and Review of Training Sequence – Slides & Demonstrations	Biographical Questionnaire, Baseline SSQ & GamePAB
	Training on OLIVE Functional Practice & Test	Usability Questionnaires SSQ Post-Practice
	Movement to Contact Exercise	
	AAR Preparation	Exercise Questionnaire, SSQ, & Short Interview w/ Leader
	AAR W/ Trainees	Recorded
	Hotwash	Exercise Controller Interview
2nd Day	Practice Exercises – Generated by Units	Ex Questionnaire, SSQ, & Leader Interview
3rd Day	Practice Exercises (AM for U.K., PM for U.S. Groups)	Exercise Questionnaire, SSQ, & Leader Interview
	Joint Operation Orders & Initial Ex	
	AAR Prep & Unit AARs	Exercise Questionnaire, SSQ, AAR Recorded
	Joint AAR	Recorded
	2nd Exercise	
	AAR Prep & Unit AARs	Unit AAR Recorded
4th Day	Practice Exercises (AM for U.K., PM for U.S. Groups)	Exercise Questionnaire, SSQ, & Leader Interview
	Joint AAR Re 2nd Exercise	AAR Recorded
	3rd Larger Exercise	
	AAR Prep & Unit AARs	Exercise Questionnaire, SSQ, & Unit AAR Recorded
	Joint AAR	Recorded

One of the major issues in planning was the military insistence that the Soldiers involved must get some training value, which meant that the data collection had to be structured around the “training” nature of the exercise. Another issue was the "coalition" framework of the exercise. Typically, interactions between coalition forces during military operations are coordinated at relatively high levels, with coalition forces conducting operations in separate sectors or relatively independent areas rather than as coordinated small or combined units in a joint mission. The non-coalition exercises (conducted at local sites) and coalition mission scenarios (linked between the U.S. and U.K.) were constructed by ex-military SMEs in order to provide believable scenario sequences for ground operations which would be conducted by a

small number of Soldiers from the two countries' militaries. A key focus was that the simulation capabilities of the OLIVE could be exercised and evaluated by the Soldiers, trainers, and observers.

The mission plans required joint convoys in an increasingly hostile environment, with the goal of escorting evacuating embassy personnel following a non-combatant evacuation order (NEO). These mission exercises were conducted while responding to conflicting information and requirements. The basic goal of the mission was to form a joint task force, move to the embassy locations and escort embassy personnel from both the U.S. and U.K. embassies from their gathering location to a pick-up location. Complications were inserted during the route to the embassy, in dealing with the embassy personnel, and during escort maneuvers.

While considerable effort was made to schedule active duty U.S. and U.K. units, only limited success was achieved. The U.S. group was recruited from West Point cadets during summer assignments, Soldiers assigned to RDECOM-STTC, and an instructor and trainees from the Captain's Career course at the U.S. Army Infantry School. The instructor served as the U.S. exercise controller/trainer and the Captains served as platoon leader and platoon sergeants. On the U.K. side, a company commander (Captain), platoon leader (Lieutenant), and a partial platoon from the 3rd Mercians Regiment provided participants for the exercise. The U.K. Soldiers were from a single platoon. While the U.K. Soldiers were from a coherent unit, and the U.S. Soldiers were not, neither participated in the exercise with the objective of meeting current training requirements.

Method

The major research issues of interest were: Graphical User Interface usability, potential training effectiveness, and support for feedback. As the groups using the GBS were from different militaries, the data gathered was compared between the groups to determine if there were cultural or programmatic differences that affected the responses to questionnaires and interviews.

Participants. In addition to the trainees described above, the scenarios also required several role players at each location, a SAF operator, and ancillary exercise control (EXCON) personnel at both locations. With the exception of the exercise control personnel, these additional personnel did not participate in data collection about the system.

The U.K. trainee group were all male enlisted Soldiers assigned by the LWC, with an average age of 21.19 (minimum = 18, maximum = 29, N = 22). The U.S. trainee group were also all male, with an average age of 26.76 (minimum of 20, maximum of 48, N = 19). The difference in age range resulted from the ad hoc nature of the U.S. group, which had Captains playing the role of squad leaders and West Point cadets as squad members. The average time in service was 2.1 years for the U.K. and 4.6 years for the U.S., although there was a considerable range for the U.S. group, again because of the mix of officers and cadets.

Materials. The computers at all locations were considered "high-end" machines, at the time, in terms of memory (generally with a minimum of two gigabytes of random access

memory) and graphics cards (all with at least 256 megabytes of dedicated graphics memory) operating above two gigahertz in processing speed. All had access to the internet and used the Windows XP^{tm} operating system. The access to the internet was required as the OLIVE servers maintained by Forterra Systems, Inc., were in California.

The questionnaires were presented and data were collected using stand-alone software on each respondent's computer. Several questionnaires were administered repeatedly during the several days of training and exercises. The planned schedule for the administration of the questionnaires is detailed in Table 1, above, and was followed to the greatest extent possible. The biographical questionnaire was used to collect basic personal information; the GamePAB and GEM collected information on game skills, experience, and knowledge; the GUI, Exercise, and AAR questionnaires were used to gather information about the GBS based on different scenarios and with different experience levels on the GBS; and the SSQ was administered after every GBS interaction as a monitor on participant's health.

Procedures. The general process was to provide familiarization training on the OLIVE system, conduct local (not multi-national) exercises that would provide both some training value and familiarization with the capabilities of the system, then conduct the series of multi-national exercises and AARs. The general sequence of activities followed the plan presented in Table 1, modified to meet technical problems and a dynamic training situation. The number of dynamic objects and operations (e.g. moving, operating equipment) in the simulation were reduced due to increasingly low visual frame rates. In addition, the AAR system was not working correctly, and therefore the AARs were conducted "in-world" using screenshots captured during exercise activities projected on a common screen (a new feature implemented in the OLIVE system just prior to the exercise). Finally, the Leaders and Trainers were quite comfortable dropping entire planned vignettes if the time required for a prior vignette ran over, or if they perceived some value to extending or enhancing a vignette (e.g. by allowing a firefight). This propensity to change the nature and sequence of operations "on the fly" led to serious disruptions in the data gathering sequences.

Following the initial familiarization with the OLIVE system, and initial questionnaires, the U.K. and U.S. groups created local exercises "on the fly" that exercised the key functionality introduced during the initial training. Both groups practiced patrol movements, movement to contact, react to contact, and some portion of checkpoint setup and operations. Following each of these exercises the Exercise Questionnaire was administered to the U.K. group, but not the U.S. group, resulting in multiple but non-matching administrations to the two groups. The Exercise Questionnaire was also administered following the planned joint exercises.

Several planned data collection procedures were altered for various reasons during the week of training and exercises. As the U.K. Soldiers did not use and were not experienced with U.S. standard AAR processes, they did not complete the AAR questionnaire. The U.S. group was more experienced in both the application and preparation of AARs and completed the AAR questionnaire. However, the OLIVE AAR did not work as planned for either the local or the joint exercises. Because of the large amount of data being recorded for the AAR, the server became overloaded causing unexpected crashes of the entire system. The work-around used for the first set of joint exercises was to provide "snapshots" of the exercise for the AAR. The local

exercises did have some replay capabilities from the OLIVE system, which were used and referenced prior to the AAR questionnaire administration.

Results

Biographical questionnaire. The general demographic factors that might influence the CMEX I results focused on general familiarity, experience, and skill with computers and software. The available responses were categorical, and are presented in Table 2, below. Both the U.S. and U.K. Soldiers started using computers at a young age, although a few U.S. Soldiers indicated that their earliest use was at 18-20 and 24-29. Almost all of the U.K. Soldiers used a computer at home (almost two-thirds owned a computer), although only three indicated using computers on the job. All of the responding U.S. group (17) used computers at home (and owned computers), and ten used computers at work. Most of the U.K. Soldiers (12) reported using icon-based programs more frequently than once a month, used menu interfaces more frequently than once a month (12), and email (14) or the internet (14) more frequently than once a month. Some caution has to be noted in the U.K. responses, in that several responders reported “never” in all of these areas, including one who claimed over twenty hours of game play a week. Most of the U.S. group (13) also reported a high frequency of icon-based program use, sixteen reported using menu-based programs more frequently than once a month, while all reported using email daily and all but one also used the internet daily.

Some of the demographic information addressed game-based experience and self-rated skills in more detail. Twelve of the seventeen U.K. Soldiers reported playing games at least monthly. Five rated video games as “a lot of fun” while nine others provided the median reply of “average enjoyment.” Five also rated themselves as “good”, four as “better than average”, and only one self-rated as “bad.” None of the U.K. Soldiers had experienced a U.S. Army game simulation. The most popular game played by the U.K. Soldiers was “Call of Duty”. In the U.S. group (a much more diverse collection) eleven of seventeen reported playing games at least monthly. Nine of the U.S. group rated video games at the top of the fun scale, with seven rating at “average enjoyment”. Only one member of the U.S. group self-rated as “good” while fifteen rated themselves as “better than average” or “average”. As might be expected, ten of the U.S. group had used “America’s Army”, and one had used “Ambush”. Thirteen members of the U.S. group had used “Call of Duty” and “Medal of Honor”.

Game performance assessment battery. The GamePAB system has several measures generated from the assessment battery, as described above. The outcomes and significant comparisons from the assessment battery are presented in Table 3, below. As indicated in the table, the Average Posture Reaction Time, Average Communication Time, Percent Aim Time on Target, and Average Shot Reaction Time did not present significant differences between the participant groups (after adjusting the significance level for the number of comparisons made with all GamePAB measures). The two significant differences between the U.K. and U.S. Soldiers’ performance occurred during a single task in which the participant is required to follow

Table 2. Biographical Questionnaire Responses on Computer Use from CMEX I

Where do you currently use a computer?	U.S. Soldiers	U.K. Soldiers
Home, Barracks, or BOQ	Yes (17)	Yes (15)
Unit Work Site	Yes (10)	Yes (3)
Do you own a personal computer?	Yes (17)	Yes (10)
Average hours per week you use a computer?	28.47 hrs.	12.0 hrs.

When did you start using computers?							
Years Old	0-5	6-11	12-14	15-17	18-20	21-23	24-29
US Soldiers	2	7	4	2	1		1
UK Soldiers	2	8	4	2	1		

How often do you use icon-based programs or software?					
	Never	Less than	Monthly	Weekly	Daily
US Soldiers	1	3	3	2	8
UK Soldiers	2	3	3	4	5

How often do you use programs or software with pull-down menus?					
	Never	Less than	Monthly	Weekly	Daily
US Soldiers		1	1	1	14
UK Soldiers	3	2	3	5	4

How often do you use email (at home or work)?					
	Never	Less than	Monthly	Weekly	Daily
US Soldiers					17
UK Soldiers	3		1	9	4

How often do you use the internet (not including email or gaming)?					
	Never	Less than	Monthly	Weekly	Daily
US Soldiers				1	16
UK Soldiers	1	2	1	9	4

What is your level of computer expertise?					
	Novice	Good w/ 1 Program	Good w/ Several	Program w/ Several	Expert
US Soldiers	2	3	10	2	
UK Soldiers	5	3	7	2	

How often do you play computer games?					
	Never	Less than	Monthly	Weekly	Daily
US Soldiers	1	1	5	9	1
UK Soldiers	3	3	1	7	3

How much do you enjoy playing video games (home or arcade)?					
	Not Very Much	Somewhat	Average Enjoyment	Lots of Fun	Most Fun in Life
US Soldiers	1		7	9	
UK Soldiers	1	2	9	5	

Please rate your skill at playing video games.					
	Bad	Poor	Average	Better than Avg.	Good
US Soldiers		1	7	8	1
UK Soldiers	1		7	4	5

and mimic the behaviors of a lead Soldier while also responding to questions about the lead Soldier's equipment and elements of the surrounding environment. The U.S. Soldiers were significantly better in both following (Percent Time Following) and responding to questions (Percentage of Correct Communications).

Table 3. GamePAB Outcomes for Soldiers in CMEX I

Measure	U.S.			U.K.			Significance*
	N	M	SD	N	M	SD	
Percent Time Following	18	36.12%	16.09	17	20.38%	11.18	$t = -3.342$ $p < .002$
Percentage of Correct Communications	16	93.23%	12.63	13	66.28%	30.96	$t = -3.182$, $p < .004$
Average Posture Reaction Time	18	1.53 s	.288	17	1.63 s	.175	*ns*
Average Communication Reaction Time	16	3.14 s	.626	10	2.86 s	2.05	*ns*
Percent Aim Time on Target	18	70.25%	14.05	17	68.18%	11.56	*ns*
Average Shot Reaction Time	18	.857 s	.5093	17	.889 s	.289	*ns*

* Significance levels were adjusted for the number of comparisons made.

Game experience measure. The GEM Questionnaire addressed game experience and preferences in greater detail than the Biographical Questionnaire. The Game Experience scale did not show any significant differences between the U.K. participants and the U.S. participants (as shown in Table 4), indicating that at least for self-reported experience there were no significant differences between the culturally different groups. The Video Game Knowledge scale did find a significant difference between the groups (see Table 4), indicating that the U.K. Soldiers had significantly less knowledge about the "popular" games than the U.S. Soldiers.

Table 4. Game Experience Measure Outcomes for Soldiers in CMEX I

Measure	U.S.			U.K.			Significance*
	N	M	SD	N	M	SD	
Video Game Experience	16	2.65	.57	17	2.46	.61	*ns*
Video Game Knowledge (Average Percentage Correct)	16	72.92	14.83	17	49.57	13.26	$t = -4.771$, $p < .001$

* Significance levels were adjusted for the number of comparisons made.

Simulator sickness questionnaire. The Simulator Sickness questionnaire (Kennedy, et al., 1992) was administered on a repeated basis to both the U.S. and U.K. participants following the initial training episodes and exercises. The use of the questionnaire was based on concerns about the potential for debilitating effects from long-term exercises on computers. This

generated eleven administrations of the questionnaire. There were minimal changes noted, with insufficient variation for any group or individual showing changes over repeated trials, any and further analysis was deemed unnecessary. None of the participants indicated any troubling symptoms developing from interaction with the simulation during the course of the exercise.

Graphical user interface questionnaire. As noted in the introduction, the individual questions used in the questionnaire are in Appendix C, which also has the anchors for the response scales. Analyses of the clustered questions was performed for each and then compared between the two groups of Soldiers. The scales were generated by calculating the mean response for the questions on the Likert response range (reversing those scales that ran from high to low so that higher numbers are always more positive). Those scales that differed in response range (seven rather than five response items) were weighted before being included in the averaging formula. (For example, responses from a seven point response scale were multiplied by 5/7 before including that response.) The scale means for each group are presented in Table 5.

Table 5. Graphical User Interface Questionnaire Scale Outcomes for Soldiers in CMEX I

Measure	U.S.			U.K.			Reliability
	N	M	SE	N	M	SE	*Combined Data*
Fidelity Scale	18	2.94	.12	16	3.01	.13	α=.839 N=31, 21 items
Control Operations Scale	18	3.38	.108	16	3.39	.14	α=.823 N=34, 11 items
Avatar Capability Scale	18	2.95	.166	15	2.95	.17	α=.853 N=33, 10 items

As can be seen in Table 5, there were no significant differences on any of the GUI scales. Overall, the ratings for all the scales were in the middle of the five point Likert response range. The exception is the Control Operations scale, which produced an overall mean of 3.388 (SE = .086, N = 34). The Cronbachs Alphas for the scales are provided in Table 5, under the Scale name.

Exercise questionnaire. As noted in the introduction, the individual questions from the questionnaire are in Appendix D, which also has the anchors for the response scales. The items included in the scales are listed in Appendix D. The scales were generated by calculating the mean response for the questions on the Likert scales (reversing those scales that ran from high to low so that higher numbers are always more positive). Those scales that differed in response range (seven rather than five response items) were adjusted by weighting (multiplied by 5/7) before being included in the averaging formula. (For example, responses from a seven point response scale were multiplied by 5/7 before including the response.) Analyses of the grouped questions was performed for each contingent and compared between the two sets of Soldiers. The means for each group on these scales are presented in Table 6, below.

As can be seen, there was no significant difference between the groups on the Fidelity Scale. The Fidelity Scale overall mean was 2.875, with the standard error for the mean being .094. Overall, the Cronbach's Alpha for the Fidelity Scale (with 18 items) was .871 over 22

complete response sets (data were not included if any individual response was missing from the participants' entire set, reducing the total response sets considered from 33 to 22). As noted above, this.

Table 6. Exercise Questionnaire Scale Outcomes for Soldiers in CMEX I

Measure	U.S.			U.K.			Significance
	N	M	SE	N	M	SE	
Fidelity Scale	18	2.93	.12	15	2.81	.15	*ns*
Training Effectiveness Scale	18	3.29	.10	15	2.98	.096	t=2.163 p < .038

As shown in Table 6, there was a significant difference in the group responses to the Training Effectiveness scale, with the U.S. perceiving significantly greater training effectiveness in the GBS exercises. For the Training Effectiveness scale, Cronbach's Alpha analysis had too few complete sets of item responses in the two groups to be considered diagnostic (U.S. = 6 and U.K. = 11).

AAR questionnaire. Because the U.K. leaders contended that the Soldiers were not experienced in U.S. style AARs, only the U.S. group actually completed the AAR Questionnaire. As with the Exercise Questionnaire scales, the AAR Questionnaire scales were generated by calculating the mean response for the questions on the Likert scales (reversing those scales that ran from high to low so that higher numbers are always more positive). Those scales that differed in response range (seven rather than five response items) were weighted before being included in the averaging formula. (For example, responses from a seven point response scale were multiplied by 5/7 before including the response.)

The U.S. Soldiers responses to the AAR Interface Capability question was a mean response of 3.09 (N = 18, SE = .095). The Cronbach's Alpha for this group of questions equaled .738 over 18 response sets, with 14 items in the scale (see Appendix F for the scale items). The responses to the AAR Training Capability questions generated a mean of 3.17 (N = 18, SE = .123). Cronbach's Alpha for AAR Training Capability was .818 over 18 complete response sets, with 10 items in the scale (see Appendix F for the scale items). Several questions required direct comparisons or evaluations and are presented in Table 7. The general consensus from the direct questions indicates that the displaying events, ease of review, and focus for future exercises were somewhat better with this system. While there were some negatives about the time to prepare and ease of preparation for the AARs, this may have reflected the difficulties in getting the system to record and replay any sequences from the missions.

Interviews and discussions. Interviews with the U.S. Cadets and U.K. Soldiers about the exercise addressed the same topics as those in the Exercise Questionnaire and the AAR Questionnaire, within an open-ended structure. In general, the coalition mission goal or context that was used for the exercise was seen as very unrealistic on several fronts. Both the U.S. and U.K. Leadership were of the opinion that the mission would not have been assigned to such a small groups of "regulars." In addition, they did not believe that a coalition commander would

Table 7. AAR Presentation Questions from CMEX I

8. How do the AAR Capabilities compare to a field training exercise AAR in the following areas?	Much Worse	Worse	Neither	Better	Much Better
a. Presentation of tasks	1	3	11	4	
b. Ability to display events	1	3	4	9	2
c. Time required to conduct exercise AAR	1	5	8	4	
d. Ease of preparation for AAR	1	5	8	5	
	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
16. The AAR system made it easy to review and determine what happened in the simulation during the exercise.		3	5	10	
17. The AAR system made it easier to determine which areas to focus upon during future exercises		2	6	9	1
(A seven point scale, only two rating at “incapable” & none at “one task”)	Few Tasks	Basic Tasks	Many Tasks	Most Tasks	All Tasks
14. In general, could this AAR support Army training as it works right now?	5	4	5	2	

lead joint coalition AARs after such exercises, as was imposed upon them by the need to test the distributed AAR functionality. Further, both groups believed that having everyone interact with civilians but without interpreters or guides diminished the realism and impaired acceptance of the effort as a viable rehearsal.

However within this negative context, the training value was seen as greatest for the leaders, and only good for introducing unit SOPs and TTPs to inexperienced Soldiers. The best aspects of use were regarded as the ability to practice the leadership and information processing skills needed during operations (although no references were made to military decision making processes). These points were emphasized by the feedback on the system features (best vs. worst), and evaluations on the best aspects of the exercises. The sounds, communications, and visual details that required leaders to continually deal with an evolving situation and control the Soldiers were the good capabilities considered to support these conclusions. In addition, the general expressed opinion was that there was insufficient environmental or operational stress supported by the system for lower level Soldier task rehearsals. That general opinion was supported by the perceived inability of the system to simulate required equipment and environment complexity. These evaluations led to the conclusion that the system could not provide any task or integrative training for Soldiers below leadership level.

Discussion of CMEX I

The first exercise was primarily intended to establish protocol and capabilities for conducting an extensive and widely distributed examination of a GBS. The goal was to conduct the exercises with Soldiers from Coalition Armed Forces, in spite of the time differential that complicated all logistical efforts. The exercise was conducted with three distributed locations over the commercial internet (the third site being occasional observers at different times and from differing remote locations), and with several local exercises conducted independently at both sites. The overall success of this effort provided considerable logistical and engineering lessons learned for the continuance of the planned series.

Framing the results. Some caveats about the exercise situations and personnel attitudes have to be introduced to frame the data collected for discussion. Both groups, despite requests from the researchers and their leadership, continually treated the exercises and data collection as less than a serious exercise. Examples of this phenomenon come from a few demographic responses that claimed the respondents never used common computer interface conventions (e.g., mouse, menus, and icons). In addition, the low-level, joint-mission framework was perceived by the unit leaders as constraining the exercise events, and that attitude inherently limited any training information that could be collected in conjunction with the exercises. The composition of the U.S. group (West Point Cadets, several Captains from the Captains Career Course at Ft. Benning, and a Major who is an instructor in that course) also meant that there was no prior unit cohesion on the U.S. side. The composition of the U.S. group also precluded establishing prior information on skills and knowledge possessed by the trainee/participants on specific tasks or drills encompassed in the local or distributed exercises. While there was a somewhat more coherent cadre from the U.K., even for the local exercises there was an observed reluctance and reticence in addressing the exercises as meaningful – in spite of the leadership emphasis in the coordination meetings on the importance of insuring that training be the primary focus.

The pre-planned and detailed exercise was designed to exercise integrated platoons in an increasingly disordered and hostile set of scenarios. The problems experienced with the software first cut down the equipment available for use and then reduced the “clutter” in the form of semi-automated forces and civilians present in the environment. By the end of the exercises, it was clear that any avatar found in the environment had a human behind it, and could become a threat at any time. As a result, the response to encounters with civilians became increasingly kinetic. This decreased the observable troop leading procedures and decision-making that had been expected, while emphasizing the initially limited kinetic aspects of the simulation. As noted in the introductory description of the GBS, it was not based in first person shooter software and therefore had limited weapons and weapons effects.

Finally, the prototype software, limited computer systems, and time constraints at the U.K. site also limited the data collection possibilities during this exercise. The U.K. experienced problems in establishing their experiment site, and moved to another site that had sufficient bandwidth for the internet-based interactions. As a result, the computers used were extremely limited in capabilities and this constrained the presentation of the software capabilities (and may have exacerbated the less-than-serious approach taken by assigned personnel). The change in location also constrained the U.K. personnel, as they were at the end of a 90-minute commute

each way on military transport. This factor led to early cessation of data collection efforts every day at that site. The information collected has been presented with the two groups separated in order to both find any differences stemming from the different organizations, equipment capabilities, and military approach as well as demonstrating the similarity of the two groups.

Information available from CMEX I. The demographics of the U.S. group seem to be representative of the active Army officer corps, in terms of age and education (in comparison to the Army Profile FY04, Office of Army Demographics, 2004). While there is little data about the computer use and game-playing demographics in the Army Profile, given that these were primarily students, they would seem to be "above the curve" in familiarity. The GamePAB and GEM information can serve as an initial baseline of capability as we work to investigate the reliability and validity of the other measures in future efforts. Overall, it is not clear whether there is any practical significance to the finding of a measurable and statistically significant difference between the U.S. and U.K. participants in terms of correct communications (basically answering questions during operations). The significant difference in Video Game Knowledge has little practical significance given the different cultures. It may have some differences in preferences or evaluations of the GBS.

Many items in the GUI questionnaire have been used before (Singer, et al., 2008), and the items are based on the standard computer game approach of using the keyboard and mouse within a "Windows" model. In addition, the game engine focus was on graphics that would be relatively available to consumers, using high end graphics. As a result it is not surprising that the general ratings for the Avatars, Control Operations, and Fidelity scales were all in the middle of the associated ranges. The best conclusion from these data is that the system did not dramatically impress, nor was the system poor, bad, or much worse than the offered comparisons and anchors. The reliability was calculated as an indicator because there were no differences between the groups, and there were at least more responses than items used in the analysis (Hinkin, 1998; Gliem & Gliem, 2003). In general there seemed to be reasonable reliability in the answers, with the exception of the GUI Training scale, which rated .235. The scales were combined based on apparent and obvious similarities in the content material, that rated similarly for the small sample of respondents, and therefore seem reasonable to use again, although the GUI Training scale requires further investigation.

The responses to the Exercise Questionnaire do not seem to align with the interview comments, as it reflects the relatively low regard that Soldiers expressed for the GBS as used within the planned exercise. The questionnaire Fidelity scale did have a grand mean slightly below the middle value in the standardized 5-point scale, although the reliability, while adequate (given the Cronbach's alpha value of .871), came from a small data set. It is difficult to reach a conclusion about the fidelity on this basis, although it must be pointed out that the questionnaire provided a more neutral context than the group interview (which can be dominated by individuals expressing strong opinions). It may also be the case that the mixture of backgrounds within the U.S. contingent (West Point cadets and Captains from the Ft. Benning course) affected the group discussion.

The significant difference on the Training Effectiveness Scale seems to indicate that the U.S. contingent accepted the system for training more readily than the U.K. It is hard to draw

any larger conclusion than that, as the comments made during final interviews indicated dissatisfactory performance in general for both groups. The possible factors mentioned above in connection with the Fidelity Scale from the Exercise Questionnaire may have played a role in this seeming dichotomy.

The interviews with the leadership in the created platoon (Captains acting as the platoon leader and sergeants, leading West Point cadets acting as lower enlisted Soldiers), focused on the leadership aspects and capabilities of the non-kinetic simulation (based on expectations established during exercise planning). They expressed their perceptions about the system providing reasonable training value in <u>leading</u> Soldiers through the exercise, but providing much less training value for the <u>Soldiers being led</u>. The multiplayer aspects and situational flexibility of the simulation were regarded as the best characteristics because those required leader skills application. The lack of simulated equipment and equipment operational fidelity (especially the poor localization of sound from weapons fire/explosions, unreasonable vehicle physics, and inappropriate wounding) were seen as major failings of the OLIVE GBS, in spite of the middle range estimate of training capability derived on the questionnaire. As noted above, the group discussions may have been dominated by differences in rank and experience.

Some lessons learned were acquired from the conduct of the exercise, although the short interval before cycling into a second mission preparation schedule precluded any large changes in the general approach. The questionnaires, as mentioned above, were transitioned to an online internet mode of acquisition in response to the difficulties in installing, administering, recovering, and removing the individual questionnaire administration software used. The questionnaires were minimally reviewed and revised during this process, based on miss-understandings and needed onsite clarifications.

Coalition Mission Exercise II

Shortly after the completion of the first exercise, a second very similar exercise was scheduled. This precluded any real changes in system functionality but did offer some opportunity to adjust the approach and data collection efforts. During this interval some of the leadership for the exercise changed, the planned exercises were altered, and the system training was revised. A major change between the first and second exercise was the involvement of a coherent unit from a 10th Mountain Division Brigade Combat Team. This U.S. contingent (a platoon minus, consisting of the platoon leadership plus two squads) was combat experienced and were in the preparation stages for another deployment overseas. They were accompanied by a Captain who acted as the U.S. exercise controller. The U.K. contingent was drawn from the 3rd Mercians at the LWC, as before, and matched the U.S. contingent in numbers and relative positions. The U.K. group was not a completely coherent unit, but was primarily drawn from a single platoon, augmented by extra assigned Soldiers, and comprised two sections (approximately equivalent to two squads). As in the U.S., the exercise controller was a Captain from the same unit. Overall this led to slightly greater numbers of Soldiers, and a few extra role-players, being involved in the second GBS evaluation exercise.

As before, the approach was constrained by the perceived needs of the military personnel taking part in the exercises and the engineering requirements for a large-scale network,

established by RDECOM-STTC military liaison and managers. Engineering constraints combined to limit the number of participants and the possible roles that participants could exercise. As noted above, the OLIVE system was developed to focus on the non-kinetic aspects of dismounted Soldier operations, and therefore had a limited range of weaponry available for use. In addition, the numbers at each physical location were targeted for platoon-minus (an incomplete platoon, without any heavy weapons). The available network connections, as well as number and capability of computers also contributed to the limitations on the number of participants at each location. In addition to the participants, the network was supposed to support exercise controllers, SAF computer systems, and role-players. The role-players were needed to provide situational stimuli for decision-making by the Soldiers.

As with CMEX-I, the key focus for the mission structure was that the simulation capabilities of the OLIVE GBS could be exercised and evaluated by the Soldiers, trainers, and observers. However, a major issue in planning was that the Soldiers involved must get some training value, which meant that any data collection had to be structured around the "training" nature of the exercise. The collection of subjective information was also complicated by the "coalition" framework of the exercise. Typically, according to Soldiers from both countries, interactions between coalition forces during military operations happen at relatively high levels, with coalition forces conducting operations in separate sectors or independent areas rather than as coordinated small or combined units with a joint mission (the framework for the experiments). The structure of the exercises seemed to interfere with the Soldiers perceptions of the exercises as training, and therefore seemed to hinder consideration of the system's potential for training during data collection.

Unfortunately, as with the first exercise, the intended venue in the U.K. was not ready for the exercise and equipment was hastily assembled at the same remote location as before. This led to reuse of some minimally capable computers and construction of others that were restricted in their graphics capabilities. The supporting servers remained in California, continuing the long haul nature of the connections between the U.S. and U.K. sites. Also complicating the experiment was a continuing focus on achieving any possible training for the Soldiers, an attitude that impaired the time and resources available for data collection. Finally, the U.K. did not constructively collaborate in the data collection instrument development, administration, or analysis. They were quite satisfied with the development of lessons learned for their facility development.

The local exercises were constructed and supported at an alternate location in the virtual world, separate and different from the geo-typical middle-eastern urban environment in which the coalition mission was to be conducted. Equipment and facilities were set up so that the two groups could conduct local exercises. This also enabled each group to leave equipment in place for follow-on situations without concern over interference by the other group. A variety of objects that could be added to or removed from the environment were developed, enabling some control over the computational load required by the extra objects during both the local and coalition exercises. The coalition mission was changed in the size of the terrain area, complexity of planned interactions, and number of objects that could be used in the exercise. These changes were made so that the computational load would be smaller at the start of the exercise, and could be reduced quickly and easily. The goal was to be able to reduce the computational load and still

enable the major structure and goal of the exercise to be conducted and achieved. These contingencies turned out to be quite necessary, as the number of personnel and amount of equipment still overloaded the client machines, particularly at the U.K. site.

The coalition mission was still based upon a NEO, but with the removal of a single individual via helicopter as the goal, followed with a second scenario that required establishing a controlling checkpoint at a chokepoint into the area of operations (a bridge that limited access from the rest of the urban area) and conducting a security patrol in the limited area. No convoy with large numbers of vehicles was planned or conducted, although a limited number of vehicles were initially provided as support during dismounted operations. These were removed from the U.K. contingent relatively early, as their machines could not handle the graphics loads caused by the additional active objects in their fields of view. The U.S. (with somewhat more capable graphics and processing) kept several vehicles for support, but conducted dismounted operations in both scenarios.

Method

Participants. The U.S. Soldiers and U.K. Soldiers were both drawn from individual units and were relatively coherent units. The U.S. Soldiers were from the same company, while the U.K. unit was drawn from a single platoon (two sections, approximately equivalent to two U.S. squads), with a few replacements based on leave or illness. The average age of the U.S. Soldiers was 22.76 ($N = 22$, $SD = 3.727$, one member over 30), and the average age of the U.K. Soldiers was 21.74 ($N = 19$, $SD = 3.364$, also with one member over 30). The average years on active duty was also comparable, with the U.S. Soldiers averaging 2.25 ($N = 22$, $SD = 1.98$) and the U.K. Soldiers averaging 2.79 ($N = 19$, $SD = 2.88$).

Materials. The questionnaires used during the first coalition mission exercise were also used during the second. As noted in the introduction, there were minor adaptations to the question scales and content as a result of lessons learned from the first coalition mission exercise. The changes, additions, and deletions to the individual questions and questionnaires are described in the questionnaire appendices (see Appendices B – F). The major difference in the data collection was the administration method used to collect responses. While individual questionnaire sets had been administered by stand-alone software on each respondent's computer during the first exercise, the administration for the second exercise was conducted using an online system. The system used was a newly developed internet available system called the Army Research Institute Virtual Laboratory (ARIVL) that used Government password protocols to protect the information, but enabled internet access to the developed questionnaires for remote data collection without a password. Unfortunately, only after exercise completion were several anomalies found that had led to the loss of responses to some questions in the GUI, Exercise, and AAR questionnaires.

The Biographical questionnaire (Appendix E) was used to acquire background biographical information about the prior educational and military experience and training of the participants. Two measures, GamePAB and GEM (Appendix B), were used to develop a baseline of the users' knowledge, experience, and skills with games. The SSQ (Kennedy, et al, 1992) was repeatedly administered to monitor any physiological side effects arising while

Soldiers performed tasks in the GBS. The GUI Questionnaire (Appendix C) was again used to address the system interface. The Exercise Questionnaire (Appendix D) was again administered to gather information on system fidelity and training effectiveness. The AAR Questionnaire (Appendix F) was administered at the end of all CMEX II exercises, using the online system.

Guided interviews and informal discussions were also conducted that addressed opinions about exercise preparation, actions taken and rationale for actions during the exercise, and post hoc appraisal of the system as well as the AAR functions. These were conducted separately with leadership and with the trainees (in the U.S.) after the exercises. The time available for these interviews, especially in the U.K., was constrained by the extended time course of the exercises.

Procedures. The general process was very similar to CMEX I, beginning with training on the OLIVE system, then conducting local exercises that would provide both some training value and familiarization with the capabilities of the system. Both groups then worked together in conducting the series of joint exercises and AARs. The general sequence of activities followed the plan presented in Table 1. These plans incorporated some simplifications in the simulation due to problems encountered in the first set of exercises (fewer vehicles, automated forces, etc.). The AAR recording system was still not working correctly, however short segments of the exercise could be recorded based on requests by the Captains at either location for use during AARs. The use of higher echelon officers as exercise controllers was also dropped.

Following the initial familiarization with the OLIVE system, and initial questionnaires, the U.K. and U.S. groups created local exercises for their units. These were created with more guidance about the pre-configured assets and exercise areas than was provided during CMEX I. The local sessions were also guided, so that they exercised key GBS functionality introduced during the initial training. Both groups practiced patrol movements, movement to contact, react to contact, and some portion of checkpoint setup and operations. These exercises, in addition to the planned joint exercises, were used as the basis for the Exercise Questionnaire responses.

Several planned data collection procedures were altered for various reasons during the week of training and exercises. Although the U.K. Soldiers did not allow recording of their review sessions after the local exercises, they did participate in the coalition mission AARs, but did not complete the AAR questionnaire. Only the U.K. leadership completed the AAR questionnaire, at the end of the week of exercises, after strong urging. The U.S. group was more experienced in both the application and preparation of AARs, and all Soldiers completed the AAR questionnaire, in addition to allowing recording of some AAR sessions. However, due to miscommunication, none of the leaders completed the AAR questionnaire.

CMEX II focused on two scenarios: 1) a NEO for Embassy personnel, and 2) Security Assistance following the NEO within the host country. As one of the Coalition goals for the exercise was to actually engage the two military organizations in collaborative and interactive efforts "on the ground", the exercises were scripted to require the forces to cooperate during portions of the scenarios. Discussions with the U.K. again established some contingencies in terms of terrorist and insurgency elements to be employed in the scenarios. As before, the scenarios were continually revised right up to role-player and exercise controller rehearsals, at

the end of the week before CMEX II. Each scenario was designed with events that were intended to force decision-making upon the "trainees." These events were also designed to support evaluation of the features, functions, and fidelity as well as enabling judgments about training effects.

The considerably simplified first scenario started with the Coalition forces dismounted at either side of a city area. As before, each platoon leader provided their Operations Orders (OPORDs) to their units prior to exercise initiation. The goal presented in the NEO was to "permissively" evacuate a high value person and material from an embassy in a host nation. This required both groups to maneuver independently to the embassy area, the U.K. establishing zone security while the U.S. established a secure helicopter landing zone (LZ). The U.K. extracted and handed over the high-value person and material to the U.S., contingent who conducted the individual and material to the LZ for extraction. The U.S. then passed through the U.K. zone security to set up a checkpoint at a nearby bridge. The U.K. conducted a separate presence patrol, during which information was pushed at the U.K. contingent for communication to the U.S. contingent. The events were constructed to address many of the standard tasks required of military personnel. Example tasks that were reviewed and included were: Troop Leading Procedures, Tactical Movement in Urban Area, as well as Conduct Roadblock and Checkpoint Operations (e.g., ARTEP 7-5-MTP).

Originally, plans required each scenario to be conducted three times with alternating AARs by the U.S. and U.K. leaders from their respective locations. The offset time schedules (between the U.S. and U.K.) provided time for one joint exercise per day, enabling three exercises over three days. The scenario events were scripted to start with U.S./U.K. forces operating separately but with coordinated efforts, and for the events to require increasingly collaborative efforts. The collaborative efforts were primarily in information gathering and decision-making, ramping up to combat by combining their forces for defense in place. The events and outcomes were scripted to depend on projected reasonable responses of the "trainees" to the demands and evolution of the situation.

As before, the scenarios required several role players, SAF operators, and ancillary EXCON personnel. During the exercise there were approximately sixty personnel in the distributed exercise. While there was still considerable direction imposed from the U.S. side, during the CMEX II exercises, control over the distant U.K. role-players was left to the U.K. local exercise controller. The U.K. local controller had access to the script for the role-players, but none of those personnel had significant opportunity to rehearse the roles. In addition, the scenarios were continually adjusted based upon engineering requirements and system performance limitations as well as training leader suggestions. This resulted in some confusion about activities that were to be performed upon demand, or in response to Soldier’s actions.

Results

Biographical information. As noted in the participants’ description for CMEX II, the U.S. and U.K. Soldiers were very similar in age and background. The basic information on the use of computers, level of experience, and familiarity with computer games or video games is presented in Table 8.

Table 8. Biographical Questionnaire Responses on Computer Use from CMEX II

Where do you currently use a computer?	U.S. Soldiers	U.K. Soldiers
Home, Barracks, or BOQ	Yes (19)	Yes (19)
Unit Work Site	Yes (8)	(0)
Do you own a personal computer?	Yes (20)	Yes (15)
Average hours per week you use a computer?	16.32 hrs.	8.24 hrs.

When did you start using computers?

Years Old	0-5	6-11	12-14	15-17	18-20	21-23	24-29
US Soldiers		11	9		2		
UK Soldiers	1	9	6	1		2	

How often do you use icon-based programs or software?

	Never	Less than	Monthly	Weekly	Daily
US Soldiers	5	1	1	5	10
UK Soldiers	6	2	5	3	3

How often do you use programs or software with pull-down menus?

	Never	Less than	Monthly	Weekly	Daily
US Soldiers	2	5	2	6	7
UK Soldiers	6	4	3	4	2

How often do you use email (at home or work)?

	Never	Less than	Monthly	Weekly	Daily
US Soldiers		2	1	2	17
UK Soldiers	3	1	4	6	5

How often do you use the internet (not including email or gaming)?

	Never	Less than	Monthly	Weekly	Daily
US Soldiers		1	2	2	17
UK Soldiers			4	7	8

What is your level of computer expertise?

	Novice	Good w/ 1 Program	Good w/ Several	Program w/ Several	Expert
US Soldiers	8	5	9		
UK Soldiers	6	7	6		

How often do you play computer games?

	Never	Less than	Monthly	Weekly	Daily
US Soldiers	6	3	1	3	9
UK Soldiers	2	2	5	4	6

How much do you enjoy playing video games (home or arcade)?

	Not Very Much	Somewhat	Average Enjoyment	Lots of Fun	Most Fun in Life
US Soldiers	2	1	7	9	3
UK Soldiers	2	1	6	9	1

Please rate your skill at playing video games.

	Bad	Poor	Average	Better than Avg.	Good
US Soldiers	2	4	5	5	6
UK Soldiers		2	11	2	4

The biographical information collected was the same as that collected during the CMEX I, addressing familiarity, experience, and skill with computers and software. While the age and service distributions were somewhat skewed, they were not abnormal or unexpected. The U.S. and U.K. response to playing computer games was similar with twelve of twenty-two U.S. Soldiers and ten of nineteen U.K. Soldiers reporting that they played computer games at least weekly. One-half of the U.S. Soldiers also claimed better than average skills with video games (11 of 22), while only six of nineteen U.K. Soldiers made the same claim. The U.S. Soldiers self-reported hours per week playing video games averaging 12.68 (SD = 14.9) and the U.K. Soldiers averaged 9.05 hours (SD = 10.01) in their responses. As may be apparent from the large standard deviations, these are not normal distributions, but are bi-modal, as is reflected in the on computer and video game use.

Game performance assessment battery. The GamePAB system has several measures generated from the assessment battery, as described above. The outcomes and significant comparisons from the assessment battery are presented in Table 9. As indicated in the table, one significant difference between the U.K. and U.S. Soldiers' performance occurred during a single task in which the participant is required to follow and mimic the behaviors of the lead Soldier while also responding to questions about the lead Soldier's equipment and elements of the surrounding environment. The U.S. Soldiers were significantly better in following (Percent Time Following) the lead Soldier.

Table 9. GamePAB Outcomes for Soldiers in CMEX II

Measure	U.S.			U.K.			Significance*
	N	M	SD	N	M	SD	
Percent Time Following	20	47.03%	10.27	14	28.5%	14.69	($t = -4.341$, $p < .001$)*
Average Posture Reaction Time	20	1.48 s	.1597	14	1.48 s	.2085	
Percentage of Correct Communications	19	95.61%	10.89	11	96.67%	7.45	
Average Communication Reaction Time	19	2.87 s	.0241	12	2.72 s	.2435	
Percent Aim Time on Target	20	70.04%	15.35	14	72.93%	9.46	
Average Shot Reaction Time	20	.78 s	.1422	14	.824 s	.2388	

* Significance levels were adjusted for the number of comparisons made with these data.

Game experience measure. The GEM scale data was calculated in the same manner as the data from CMEX I. The Video Game Experience scale (averaged over the Likert scale experience ratings) did not show any significant differences between the U.K. participants and the U.S. participants (as shown in Table 10), indicating that at least for self-reported experience there were no significant differences between the culturally different groups. The Video Game Knowledge scale did find a significant difference between the groups (see Table 10), indicating

that the U.K. Soldiers had significantly less knowledge about the selected "popular" (and probably U.S. centered) games than the U.S. Soldiers.

Table 10. Game Experience Measure Outcomes for Soldiers in CMEX II

Measure	U.S.			U.K.			Significance*
	N	M	SE	N	M	SE	
Video Game Experience	21	2.8	.146	19	2.83	.126	
Video Game Knowledge (Average Percentage Correct)	21	68.934	3.345	19	54.386	3.647	$t = -2.945$, $p < .005$

* Significance levels were adjusted for the number of comparisons made.

Simulator sickness questionnaire. Although the questionnaire was administered and monitored regularly throughout the several days of exercises, there were no indications of excessive change or discomfort from the use of the GBS. As with the CMEX I data, no further analyses were performed.

Graphical user interface questionnaire. As noted in the introduction, the individual questions used in the GUI Questionnaire are in Appendix C, which also has the anchors for the response scales. The scales were generated from the response data and then compared between the two groups of Soldiers. The scales were generated using the same procedures used in CMEX I, by calculating the mean response for the questions on the Likert response range (reversing those scales that ran from high to low so that higher numbers are always more positive). Those scales that differed in response range (with seven rather than five response items) were weighted before being included in the averaging formula. The means for each group on these scales are presented in Table 11, below.

Table 11. Graphical User Interface Questionnaire Scale Outcomes for Soldiers in CMEX II

Measure	U.S.			U.K.			Overall
	N	M	SE	N	M	SE	
Fidelity Scale Alpha = .876 (12 Items)	19	3.265	.125	17	3.243	.149	3.255 N = 36, SE = .095
Control Operations Scale Alpha =.871 (7 Items)	19	3.75	.132	17	3.68	.150	3.72 N = 36, SE = .098
Avatar Capability Scale Alpha = .868 (7 Items)	19	3.241	.161	17	3.244	.170	3.242 N = 36, SE = .115

As noted in the procedures, the questionnaires were transitioned to new software before CMEX II. During data analysis, we discovered that two sets of questions that were used by the different scales had very low response rates. As all other items had responses, the scales were adjusted for analysis by removing those two question sets from the scales. The resultant group and overall means are presented in Table 11, showing no significant differences on any of the adjusted GUI scales. Overall, the ratings for all the scales were in the middle or slightly above

on the five point Likert response range (when averaged). As there was a sufficient ratio of responses to items in the scales, Cronbach's alphas were calculated in a preliminary analysis of scale reliability.

Exercise questionnaire. Minor changes to the questions were made between CMEX I and II, as described in Appendix D. There were data collection software problems during CMEX II, resulting in considerable missing data. Several questions used in the scales had no data, while other questions were missed by some of the respondents. We decided not to compile scales with partial data, nor report question responses that did not have at least a 70% response rate. The mean response for items from the Fidelity Scale are presented in Table 12, including a second administration to the U.S. contingent following the last joint exercise.

Table 12. Exercise Fidelity Scale Items from CMEX II

Question Stem	Response Scale	U.S. 1st EX N = 20	U.K. 1st EX N = 14	U.S. 2nd EX N = 22
2. How much did the animated gestures contribute to this exercise?		M = 4.45 (SE = .266)	M = 5.07 (SE = .474)	M = 3.73 (SE = .373)
(1) Limited capabilities hindered activities	(7) Capabilities supported many key activities			
5. During the exercise, were there any important sounds missing?*		M = 2.55 (SE = .223)	M = 1.36 (SE = .169)	M = 2.36 (SE = .155)
(1) None	(5) All			
6. During the exercise did any important sounds seem wrong?*		M = 1.6 (SE = .184)	M = 1.29 (SE = .163)	M = 1.91 (SE = .173)
(1) None	(5) All			
8. Was there any noticeable latency in the simulation that affected the exercise?*		M = 3.05 (SE = .185)	M = 2.93 (SE = .245)	M = 2.95 (SE = .154)
(1) System was always fast enough for the exercise	(5) System was never fast enough for the exercise			
10. Did the explosions and special effects seem real enough for training in these exercises?		M = 3.3 (SE = .147)	M = 3.29 (SE = .304)	M = 3.32 (SE = .191)
(1) Too fake for any training	(5) Good, will improve Soldier performance			
11. Was the local voice system (not radios) adequate to support this training exercise?		M = 3.0 (SE = .192)	M = 3.57 (SE = .173)	M = 3.14 (SE = .190)
(1) Inadequate	(5) More than adequate for training			
18. Were the simulated radios adequate for these scenarios?		M = 3.61 (SE = .257) (N = 18)	M = 3.62 (SE = .241) (N = 13)	M = 3.74 (SE = .240) (N = 19)
(1) The radios didn't support the communications in the exercise.	(5) The radios supported the needed communications well, enabling focus on the training event.			

*Note that the scale is reversed, with one as the most positive response.

The same solution was applied to Training Effectiveness Scale items in order to generate information that could be used for inferring possible training effect. Applying the rule that at least a 70% response rate is considered reasonable resulted in information for only six questions. It should be noted that questions 10, 11, and 18 from the Fidelity Scale would also be considered part of the Training Effectiveness Scale, and are presented in Table 12. The descriptive data for the remaining questions from the Training Effectiveness Scale are presented in Table 13.

Table 13. Exercise Questionnaire Training Effectiveness Scale Items from CMEX II

Question	U.S. 1st	U.K. 1st	U.S. 2nd
3. How does the simulation compare to field training exercises in:*	Response Scale (1) Much better (5) Much worse		
a. the diversity of tasks	M = 2.93 (SE = .182) U.S. = 15	M = 2.83 (SE = .207) U.K. = 12	M = 2.81 (SE = .131) U.S. = 21
b. the ability to record events for review & analysis	M = 2.4 (SE = .163) U.S. = 15	M = 2.73 (SE = .237) U.K. = 11	M = 2.31 (SE = .133) U.S. = 13
c. the time required for exercise	M = 2.64 (SE = .199) U.S. = 14	M = 3.0 (SE = .270) U.K. = 11	M = 2.65 (SE = .191) U.S. = 17
d. the ease of change in exercise	M = 2.79 (SE = .214) U.S. = 14	M = 2.90 (SE = .233) U.K. = 10	M = 2.47 (SE = .133) U.S. = 15
16. How well did each of the following areas support working as a team to accomplish the unit's mission in this exercise?	Response Scale (1) Prevented (5) Enabled		
a. Visual aspects	M = 3.35 (SE = .221) U.S. = 20	M = 3.43 (SE = .251) U.K. = 14	M = 3.05 (SE = .158) U.S. = 22
b. Gesture	M = 3.3 (SE = .231) U.S. = 20	M = 3.29 (SE = .286) U.K. = 14	M = 3.00 (SE = .197) U.S. = 22
c. Communications aspects	M = 3.55 (SE = .223) U.S. = 20	M = 4.0 (SE = .257) U.K. = 14	M = 3.73 (SE = .188) U.S. = 22
d. Movement system	M = 3.55 (SE = .170) U.S. = 20	M = 3.36 (SE = .289) U.K. = 14	M = 3.73 (SE = .150) U.S. = 22
21. Was the simulation adequate for rehearsing or learning Escalation of Force and Rules of Engagement?* (1) The Simulation supported ALL EOF/ROE aspects or activities. (5) The Simulation did not support any EOF/ROE aspects or activities.	M = 2.45 (SE = .256) U.S. = 20	M = 2.86 (SE = .275) U.K. = 14	M = 2.18 (SE = .193) U.S. = 22

* Note that this scale is reversed, higher scores are more negative.

AAR questionnaire. The AAR Questionnaire was administered to the U.S. Soldiers, and to the platoon leader and trainer with the U.K. contingent but not the U.K. Soldiers. The questions and stems were not altered for the CMEX II administration, and are presented in Appendix F. Two scales were generated from the AAR questionnaire: an Interface Capability scale and a Training Capability scale. As before, the scales were generated by calculating the mean response for the questions on the basic Likert scale used (5 points). The scales were adjusted for different sized Likert scales (e.g. seven point scales were multiplied by 5/7), and set to reflect higher numbers as being more positive. In four related questions in which yes/no answers were acquired, the responses were summed (yes = 2 and no = 1) and subtracted from nine, which generated a response scale from one to five with five being more positive. The questions used in these scales are identified and described in Appendix F.

The U.S. Soldiers' responses to the AAR Interface Capability scale was a mean response of 3.89 (N = 22, SD = .417), with the Cronbach's alpha being .656 (N = 22, Items = 9). The AAR Training Capability scale was 3.65 (N = 22, SD = .409), with the Cronbach's alpha equaling .628 (N = 11, Items = 10). The U.K. Leaders responses (N = 2) were 4.37 for the AAR Interface Capability scale and 3.86 for the AAR Training Capability scale.

Several individual questions in the AAR Questionnaire addressed the presentation of information for review by trainees. The direct question stems and overall response numbers for those questions are presented in Table 14, using all responses gathered.

Table 14. AAR Questionnaire Presentation Questions from CMEX II

How do the AAR Capabilities compare to a field training exercise AAR in the following areas?	Much Worse	Worse	Neither	Better	Much Better
Presentation of tasks			7	12	
Ability to display events			2	10	
Time required to conduct exercise AAR		1	11	7	
Ease of preparation for AAR		3	7	7	
	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
AAR system made it easier to determine which areas to focus upon during future exercises			3	13	8
AAR system made it easy to review and determine what happened in the simulation during the exercise.	1		2	13	8
(Only one response of "incapable" & none at "one task")	Few Tasks	Basic Tasks	Many Tasks	Most Tasks	All Tasks
Could this AAR support Army training as it works right now?	2	6	4	7	4

Leader interviews. Interviews were conducted with the Leaders and Trainers after the exercise series was completed. In general, all were politely complimentary concerning the potential for the GBS in training. The largest negative was the use of the system for training unit members – the Soldiers were not perceived to have gotten much training at all, nor were they projected to benefit from the application of the system until much more functionality becomes available. A consistent point was that the functionality needed wider ranging military systems (weapons, radios, night-vision systems, etc.). According to the leaders, the best use of the current technology was in enabling small unit leaders to exercise what the U.K. refers to as judgmental training. The U.S. leadership group concurred in this estimation, maintaining that decision making could be exercised, but the military procedures framing the acquisition and use of decision processes was limited and in some cases non-supportive.

When the U.K. leadership was questioned directly about the basic tasks performed during the repeated scenarios, some differences between the sections were acknowledged. It was clear to them that one section steadily improved in movement and communication, while the other did not. No explanation was generated during the interview that accounted for the difference. They also noted that the technology did support reviewing Soldier activities for training and evaluations. The weakest aspect observed was that the GBS scenarios left some of the Soldiers un-engaged during the missions, although when this situation was discussed they admitted that this was normal even during field training.

Discussion of CMEX II

The goal of the data collection effort was to gather subjective opinion on the demonstrated or potential training usefulness of the GBS, based on experiences gained while working through the exercises, either local or coalition. In order to frame that subjective opinion, background information on computer experience and gaming was collected from each group of Soldiers. While much of that information presented by country, the intent was to show that there were only limited differences between the two groups of Soldiers. This means that the general opinion data about the training effects can be interpreted without cultural distinctions.

Biographical & background information. It is not surprising that there were few differences between two groups of Soldiers. The cultural and technological differences between the U.S. and U.K. are presumed to be relatively small. Soldiers are recruited at similar ages, and experience similar activities with computers and simulations when growing up. They begin working with computers early, primarily in school. Slightly more of the U.S. Soldiers reported owning computers, and a third of the U.S. Soldiers report using computers in the work place while none of the U.K. Soldiers reported using computers on duty.

There were few differences in experience or onset of use with the gaming information, although the U.S. Soldiers claimed better skill levels and reported playing far more hours per week. During the actual tests of proficiency using GamePAB, a small significant difference was found with the U.S. spending a greater percentage of time correctly following the programmed leader than the U.K. Soldiers. The U.S. group was significantly more knowledgeable on the games questions in the GEM than the U.K. Soldiers. It may be that the GEM, being constructed

by U.S. game-players, was biased toward popular U.S. games and did not validly tap the game knowledge of the U.K. participants.

GBS questionnaire information. The GUI Questionnaire addressed control operations, fidelity, and avatars separately from the exercise activities. The new data collection system missed a high percentage of responses for two questions, which were dropped from analyses. The GUI Control scale did not differ significantly between the groups and had a reasonable reliability. The average scale score was above the middle in the response scale, indicating no large problems and moderate satisfaction with the controls used to operate the GBS. As the OLIVE system is adapted from relatively standard personal computer game format, this is not truly surprising. There seems to be little that could be done to make the keyboard and mouse control truly remarkable and easier than it already is. The GUI Fidelity scale was also similar across groups and slightly above the middle of the response scale in ratings. Again, the reliability was reasonable. This seems to lead to the conclusion that the realism of the GBS was reasonable but not impressive. The GUI Avatar assessment presented a similar picture, with no significant differences between the groups, middle of the scale acceptance ratings for the representation and interactivity, and reasonable reliability for the scale. The interview comments concerning GBS usability focused on the fidelity needs in terms of the physics and functionality of the equipment and personnel interactions, especially equipment or interactions deemed as needed in the exercise.

The Exercise Questionnaire results for the CMEX II also had low item response rates for many items, and therefore the questions with reasonable rates were presented rather than the scale means as with CMEX I. Most of the questions on fidelity had above middle of the scale response means, and the U.S. opinions did not change dramatically between administrations. The best responses were focused on the sounds presented (also supporting training effectiveness). The responses also indicate that the U.K. Soldiers liked the gestures, sounds, and voice system somewhat more than the U.S. Soldiers did. The Training Effectiveness questions presented a similar result pattern, with most responses near the middle of the scales, relatively equivalent, and not changing dramatically for the U.S. between administrations. Overall, the responses seem to indicate that the Soldiers in both groups accepted the GBS capabilities in supporting the scenarios as being reasonably effective in allowing them to work on necessary task and team skills. Perhaps the best indicator of possible training value in using the GBS is supported by the responses to how well the system supported teamwork in accomplishing exercise goals (Table 13, #16 c & d,), with responses for the U.S. increasing on the second administration for the communication aspects and movement system. In addition, the discussions indicated that all Soldiers and the Leader/Trainers felt that the system (although not presenting a normal exercise for any of the Soldiers) was in general capable of supporting field training. They felt that they were provided with information and stimuli that mostly enabled them to work through the mission, in spite of some difficulties with the simulation system and simulated equipment.

The AAR questionnaire was intended to gather information from both the trainee's point of view, and from the perspective of the trainer using new technology to perform AARs. Unfortunately, the problems with the recording system as well as the conduct and sequences of the exercise as driven by the exercise control and trainers limited the opportunity to gather

information on the actual AAR capabilities. The data that was gathered indicates moderately good acceptance for the interface and training aspects. Both generated scales had moderately low reliability, probably due to the low number of items and limited responses. The U.K. leader and trainer generated higher scale values for the two scales, which was somewhat surprising as they had explained that they didn't perform AARs like the ones that the system was designed to support. Unfortunately, the U.K. leadership would not allow observation of their discussions about the local exercises . They also objected when pressed for comments about their AAR discussions. The support staff in the U.K. also emphasized that the issue not be pursued as the Soldiers were only available for a limited time and declined to discuss them with ARI personnel.

The interviews provided less information than was hoped, although the comments made were generally positive. The most valuable information provided focused on the need for more functionality and reality. The capacity for training leader decision-making was emphasized, but problems were seen in using a wide-ranging simulation to appropriately frame and support those decision processes. In spite of all the problems and issues, the U.K. and U.S. trainers all rated the AAR capabilities as good training tools.

Lessons learned. One issue that led to negative secondary effects on the data collection efforts was that the U.K. systems were not available for testing with the GBS software prior to the U.S. contingent arriving for exercise support and data collection. Therefore, the software was not completely tested on the final equipment configuration prior to the initiation of exercises. As a result, a considerable amount of effort went into trouble-shooting the system ahead of the presentations and data collection protocol during the week-long series of exercises. These distractions also led to lower levels of task focus on the part of the Soldiers, and to a certain amount of system scapegoating when Soldiers performed poorly (e.g., the system limited their capability to conduct security observations). The lesson learned and possible cure is that full system tests should be completed prior to any usability or effectiveness investigations, and all detected issues resolved.

Unfortunately, when the U.K. client systems and internet connections slowed and some system crashes were experienced, changes to the exercise plans were implemented with minimal coordination between the U.S. & U.K. sites. The limited coordination was not through lack of attempts at collaboration, but through limitations on the communications channels available to the control personnel. The original plan was that all exercise coordination would be conducted using the communications capabilities of the GBS. However, when the system became unstable and crashed, exercise personnel had to resort to using cell phones to communicate and coordinate. The lesson learned from this situation is that communications channels for the exercise controllers must be solid, and back-up capabilities tested and available, before system exercises begin.

Conclusions & Recommendations

Game-Based Simulation System

The OLIVE GBS provides considerable scope for general dismounted Soldier training. The system supports reasonable aspects of moving, interacting, and communicating. The GUI is

acceptable, and easily controls the functions and menus needed for interactions in the virtual world. There are sufficient physics for vehicles, environment interactions, and weapons effects that can frame the employment decisions that are critical for Soldier's needs in many basic operational tasks, although there are problems in handling the numbers required for platoon level staffing (either Soldiers or opposing forces) in the exercises. Perhaps the greatest capabilities lie in the review capabilities, which enable distributed review of trainer selected replays. The communication capabilities also support learning interactions during these reviews.

The major drawback to the use of the system seems to lie in the generality of the OLIVE virtual world. Soldiers typically focus on the most forceful aspects of their jobs, as those are inherently more dangerous to them and more critical to forcing others into compliance. The OLIVE system did not provide the wide range of equipment that the military employs, which seemed to decrease acceptance of the simulation for those non-kinetic aspects that were achieved. That is, it seemed that if much of the varied equipment needed for an organized contingent was present, the Soldier's might focus more easily on the non-kinetic and informational aspects of their normal operating environment, rather than being distracted by the lack of normally present equipment in all variations. While it is easy to set up informational interactions that drive military decision processes with the OLIVE system, the Soldiers focused more on the missing components and capabilities of their mission equipment sets rather than the portions that were available.

Evaluations

The information gathered and conclusions generated are constrained by the limited nature of the exercises that were conducted. The organizational emphasis on flexible and Soldier choice-driven training during the system evaluations further limited the amount and type of information that could be collected about the GBS characteristics and functionality. The constraints on interventions and data collection therefore limit the conclusions that can be drawn from the experiments. If the military leadership had been more involved in addressing the training aspects within the context of experimentation, it is possible that the information elicited from them may have been more diagnostic in terms of specific military tasks.

In spite of the noted constraints and GBS deficiencies, the information gathered during the two experiments demonstrates that larger scale exercises can be conducted with widely dispersed contingents. This type of GBS is usable by military personnel engaged in military activities (even if non-doctrinal). Further, the Soldiers involved generally accepted the GBS and perceived some benefit from the exercises, in spite of many functionality and equipment deficiencies in the system. Soldiers also seemed to accept that GBS could be used for training at their home stations. Questionnaire responses indicated that Soldiers found the system to be easier to work with than the more logistically difficult real-world training and rehearsal activities. Any GBS used as an adjunct to home station exercises requires much more complete and tailorable set of military equipment, as well as the capability to handle the larger graphics loads imposed by more personnel using that equipment.

The most effective approach to a training experiment requires early involvement with trainers who understand both the training needs of the U. S. Army <u>and</u> the capabilities of the

equipment that will be used. The Soldier/trainers need to work in concert with the experimenter/evaluators to structure events so that system capabilities are appropriately applied in ways that can be measured. This collaborative focus can provide information about the range of GBS capabilities needed for wide-ranging training of Soldiers before an exercise/experiment is even conducted. Applying GBS to address known training requirements in the context of expert evaluation will also enable clear information on the capability of the GBS to effect needed improvement in Soldier skills, knowledge, and performance. That information can then be used to determine the efficacy of adding a general GBS to the training arsenal of the U.S. military.

Future Efforts

The RDECOM-STTC METER program for GBS is working toward greater involvement with other military simulations, more international partners, and larger networks. The goal is to provide information for future coalition training and mission rehearsal efforts in conjunction with the wide range of future coalition efforts that might arise. The approach will still focus on lower level interactions (below company level) in the context of a wider operational environment. The next incremental step is planned to involve further testing of the simulation center being implemented by the United Kingdom, and the addition of other GBS.

GBS Measurement. A major goal in these efforts for ARI has been developing, testing, and evaluating different measures and protocols that can be used in evaluating the critical aspects of different GBS systems. The intent is not to directly compare GBS in a competitive framework, but to be able to establish common measures of functionality, characteristics, and capabilities that can be easily applied. This is not a trivial or easily achievable goal.

One aspect of investigation that will continue to be pursued with GBS systems is the complex of knowledge and skills that the trainee brings with them to the learning situation. In using GBS for training and rehearsal, the intervening interface and expectations of operation can support positive or negative transfer. The level and amount of system training that first has to be accomplished may make significant impacts on the training time available. The amount of information and needed practice may also require extra preparation, or if the system follows standard conventions there may be little need for in-depth training.

The biographical information that has been collected and reported here is similar to data that has been collected and used in other work (Singer, et al., 2008). The biographical questionnaire stems from efforts that were started in the last decade (Fober, et al., 2001). The biographical data is collected and presented in order to identify the range of knowledge and skills possessed by typical users (especially Soldiers), as the generational trends continue to increase the amount of digital knowledge and skill used to employ electronics in everyday life (Singer, et al., 2008). The clear trends indicate increasing digital literacy in the Soldier population, which will certainly affect the amount and type of training required for any digital system.

In addition, the GamePAB and GEM will also need to be improved, validated, and eventually updated for the same reasons. The usability of the GBS interface is also a major factor in the assessment of GBS systems, and will continue to be used as new capabilities and

equipment are simulated, or new interface functionality is developed. All of these instruments will be applied in the next effort, described below.

Another key aspect addressing the assessment of GBS will continue to be the fidelity represented in the virtual world. Aspects of reality are required for acceptance of the environment. User acceptance of simulation has typically been based on their perception that specific aspects of reality are adequately replicated in the environment. That user acceptance of the environment as one in which “real” training can occur is the necessary foundation of any simulation-based training approach. The next important aspect of representation is that the equipment and operations that are key to the training objectives are presented with adequate realism for training transfer to occur. As always, this requires a thorough understanding of the many factors which enable learning the specific knowledge or skills. The application of trainer and leader questionnaires and interviews as the initial approach for addressing these fidelity and acceptance issues has typically been the most practical approach, and will continue to be improved and tested.

A third area of investigation concerns the instructional methods and tools that are or can be used to improve the effectiveness of GBS. In general, GBS provides an opportunity to practice performing tasks and to receive feedback on that performance that can be used to bridge the gap between straight-forward information presentation or familiarization (e.g., didactic instruction) and more realistic real-world activities (e.g. field training exercises) that are used to provide practice and certify readiness to perform acceptably. However, GBS technologies were not developed for training purposes, and the Army lacks both experience in using GBS within a training program as well as research-based training methods for using GBS in training. In addition, the use of GBS systems requires aids for scenario development, training practices, and performance measurement tools that do not exist. Training distributed teams presents additional training and performance measurement challenges in the use of GBS technologies to address Army training needs. In addition, any GBS should include, or be connected to, an AAR system. The features, functions, and capabilities of the applied AAR systems also have to be categorized and measures of applicability and effectiveness applied. These issues will be investigated in future planned events.

Coalition Mission Exercise III. The next effort is scheduled for November ’09, and will again involve Soldiers from the U.S. and U.K. Major differences in the distributed network will be tested, as the Land Warfare Center in the U.K. is finishing their equipment configuration for ongoing GBS research and development. In addition, RDECOM-STTC is hosting the required servers onsite, rather than using commercial servers in California. Equipment is being upgraded at all sites, in efforts to reduce the constraints on active objects and numbers of participants. Current plans include greater levels of instruction on the GBS systems, including more detailed and structured local exercises. Plans currently call for the leadership of the U.S. platoon to be brought into the planning cycle much earlier, and information about Soldiers capabilities and needs are being more thoroughly considered. Finally, an additional GBS is being included in the experimental exercises. This will require two separate coalition mission sessions, one with each of the systems. While the intent is not to construct a “head to head” competition, considerable effort is being made to include the widest range of task elements possible in order to address all of the strengths and weaknesses of the systems.

References

ARTEP 7-5-MTP: Mission Training Plan for the Stryker Brigade Combat Team Infantry Rifle Platoon and Squad, Army Training and Evaluation Program 7-5-MTP. (2003). Headquarters, Department of the Army, Washington DC.

Bohemia Interactive (2009). U.S. Army Selects VBS2 for Enterprise License. (press release dated 5 Jan. 2009), retrieved from http://virtualbattlespace.vbs2.com/index.php

Chertoff, D., Jerome, C., Martin, G., & Knerr, B. (2008). GamePAB: A game-based performance assessment battery application. *Proceedings from the 52nd Annual Meeting of the Human Factors and Ergonomics Society Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

FM 7-0: Training the Force. (2002). Headquarters, Department of the Army, Washington, DC.

Fober, G. W., Bredthauer, J. L., & Dyer, J. L., (2001). Computer background of Soldiers in FORSCOM units: FY00. (RR-1778, ADA399393) Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. Presented at the 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education, retrieved April 2010 from https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem+&+Gliem.pdf

Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. Retrieved 19 April 2010 from http://pcbfaculty.ou.edu/classfiles/MGT%206973%20Seminar%20in%20Research%20Methods/MGT%206973%20Res%20Methods%20Spr%202007/Week%207/Hinkin%201998%20ORM%20A%20brief%20tutorial%20on%20the%20development%20of%20measures.pdf

Kennedy, R. S., Lane, N. E., Lilienthal, M. G., Berbaum, K. S., & Hettinger, L. J. (1992, Summer). Profile analysis of simulator sickness symptoms: Application to virtual environment systems. *Presence*, 1(3), 295-301.

Office of Army Demographics, (n.d.) *Army Profile FY04*, Department of Army, retrieved July 22, 2009 from http:www.army.mil/references/FY04ArmyProfile.pdf

Office of the Undersecretary of Defense for Acquisition, Technology and Logistics (2003). *Defense Science Board Task Force on Training for Future Conflicts.* (ADA429010). Washington, DC.

Robson, S. (Nov, 2008). Not playing around: Army to invest $50M in combat training games. Stars & Stripes, retrieved 29 Jan. 09 from: http://www.stripes.com/article.asp?section=104&article=59009

Singer, M. J., Long, R., Stahl, J., & Kusumoto, L. (April, 2008). Formative Evaluation of a Massively Multi-Player Persistent Environment for Asymmetric Warfare Exercises. (Technical Report 1227, ADA 480016). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Stanney, K. M., & Kennedy, R. S., (1997). The psychometrics of cybersickness. *Communications of the ACM*, 40(8), 66-68.

Taylor, G., Singer, M. J., & Jerome C. (October, 2009). Development and Evaluation of the Game-Based Performance Assessment Battery (GamePAB) and Game Experience Measure. *Proceedings of the Human Factors & Ergonomics Society 53rd Annual Meeting,* San Antonion, TX. Human Factors & Ergonomics Society, pp. 2014-2018.

Appendix A. List of Acronyms.

AAR	After Action Review
ARI	U.S. Army Research Institute for the Behavioral and Social Sciences
ARIVL	Army Research Institute Virtual Laboratory
CAS	Close Air Support
CMEX-I	Coalition Mission Experiment One
CMEX-II	Coalition Mission Experiment Two
COE	Current Operating Environment
CTC	Combat Training Center
DI	Dismounted Infantry
EST2000	Engagement Skills Trainer (2000)
EXCON	Exercise Controller
GamePAB	Game Performance Assessment Battery
GBS	Game-Based Simulation
GEM	Game Experience Measure
GUI	Graphical User Interface
LWC	Land Warfare Centre
LZ	Landing Zone
METER	Multinational Experimentation for Training, Evaluation and Research
MOUT	Military Operations in Urban Terrain
MMOG	Massively Multiplayer Online Game
NEO	Non-combatant Evacuation Order
OLIVE	OnLine Interactive Virtual Environment
OPORDS	Operations Orders
RDECOM-STTC	Research, Development, and Engineering Command – Simulation and Training Technology Center
SAF	Semi-Automated Forces
SME	Subject Matter Expert
SOP	Standard Operating Procedures
SSQ	Simulator Sickness Questionnaire
STX	Situational Training Exercise
TTCP	The Technical Cooperation Panel
TTP	Tactics, Techniques, and Procedures
VOIP	Voice Over Internet Protocol

Appendix B. Gaming Experience Measure

Answer the questions below to characterize your previous experience with video and computer games. For each question select the appropriate choice that most accurately describes your experience. Please consider all five choices in making your selection, including those that do not have descriptive labels. Answer questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.

Participant Number ___________

		Low		Average		High
1.	What is your level of confidence with video games in general?	O	O	O	O	O

Hours per week

2. How many hours per week do you currently play video games?
   - O 0-9 hours
   - O 10-19 hours
   - O 20-29 hours
   - O 30-39 hours
   - O 40+ hours

Hours per week

3. What is the maximum number of hours per week you've ever played?
   - O 0-9 hours
   - O 10-19 hours
   - O 20-29 hours
   - O 30-39 hours
   - O 40+ hours

Number of times

4. About how many times have you read a video game magazine or website to find out tips to improve your gaming skill?
   - O 0-9 times
   - O 10-19 times
   - O 20-29 times
   - O 30-39 times
   - O 40+ times

5. How often do you play:

	Never	Rarely	Monthly	Weekly	Daily
Adventure - Graphical (e.g., Myst, Fable)	○	○	○	○	○
Adventure - Text-based (e.g., ZORK)	○	○	○	○	○
Puzzle (e.g., Minesweeper, Tetris)	○	○	○	○	○
Racing (e.g., Need for Speed, Test Drive)	○	○	○	○	○
Role-playing (e.g., Final Fantasy)	○	○	○	○	○
Simulation (e.g., Flight Simulator, Trains)	○	○	○	○	○
Sports (e.g., Madden Football, FIFA Soccer)	○	○	○	○	○
Strategy Real-time (e.g., Age of Empires)	○	○	○	○	○
Strategy - Turn-Based (e.g., X-Com: Apocalypse)	○	○	○	○	○
First Person Shooter (e.g., Half-Life, Unreal)	○	○	○	○	○
Multiplayer (e.g., World of Warcraft)	○	○	○	○	○
Online (any of the above titles in online mode)	○	○	○	○	○

6. List your recent favorite 5 game titles in the blanks.

A ______________________________
B ______________________________
C ______________________________
D ______________________________
E ______________________________

7. Indicate your experience with each game you listed in question 6 above.

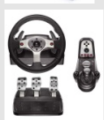
	None	Very Little	Average	High	Expert
A	○	○	○	○	○
B	○	○	○	○	○
C	○	○	○	○	○
D	○	○	○	○	○
E	○	○	○	○	○

8. Indicate your experience with the following types of game controllers:

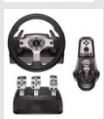
	None	Very Little	Average	High	Expert
A. 	○	○	○	○	○
 B.	○	○	○	○	○
C. 	○	○	○	○	○
 D.	○	○	○	○	○
E. 	○	○	○	○	○
F. 	○	○	○	○	○
G. 	○	○	○	○	○
H. 	○	○	○	○	○

For the following question, look at the accompanying screenshots of video games and answer the questions for each:

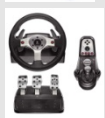
9.

	A.	B.	C.	D.	E.	F.	G.	H.
Which controller from question 8 above would you most likely use with this game? A.  B.  C.  D.  E.  F.  G.  H. 	○	○	○	○	○	○	○	○
If you were controlling the character on the right, what controller actions would you perform to defeat the enemy (button press, joystick movement, etc.)? A. Right mouse button click B. ‘A’ button press C. ‘X’ button press D. Red button press E. Spacebar	○	○	○	○	○	○	○	○
Would your enemy most likely be controlled by the computer or another person? A. Computer B. Person C. Either Computer or Person	○	○	○	○	○	○	○	○

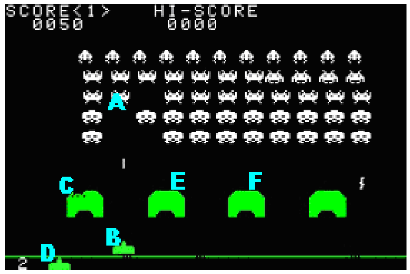
10.

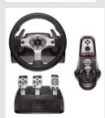
	A.	B.	C.	D.	E.	F.	G.	H.
Which controller from question 8 above would you most likely use with this game? A.  B.  C.  D.  E.  F.  G.  H. 	○	○	○	○	○	○	○	○
If you were controlling the character facing you, what controller actions would you perform to defeat the enemies? A. Right mouse button click B. ‘B’ button press C. ‘X’ button press D. Red button press E. Spacebar	○	○	○	○	○	○	○	○
Would your enemy most likely be controlled by the computer or another person? A. Computer B. Person C. Either Computer or Person	○	○	○	○	○	○	○	○
Which enemy are you currently attacking, the one on the left of the screen or the right? A. Left B. Right	○	○	○	○	○	○	○	○

11.

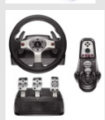
	A.	B.	C.	D.	E.	F.	G.	H.
Which controller from question 8 above would you most likely use with this game? A.  B.  C.  D.  E.  F.  G.  H. 	○	○	○	○	○	○	○	○
If you were controlling the character on the left, what controller actions would you perform to defeat the enemies? A. Right mouse button click B. 'B' button press C. 'X' button press D. Red button press E. Spacebar	○	○	○	○	○	○	○	○
Would your enemy most likely be controlled by the computer or another person? A. Computer B. Person C. Either Computer or Person	○	○	○	○	○	○	○	○
Which enemy are you currently attacking, the one on the left of the screen or the right? A. Left B. Right C. Neither	○	○	○	○	○	○	○	○

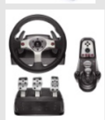
12.



<table>
<tr><th></th><th>A.</th><th>B.</th><th>C.</th><th>D.</th><th>E.</th><th>F.</th><th>G.</th><th>H.</th></tr>
<tr><td>Which controller from question 8 above would you most likely use with this game?
A.  B. 
C.  D. 
E.  F. 
G.  H. </td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td></tr>
<tr><td>If you were controlling the character marked with the letter B, what controller actions would you perform to defeat the enemy?
A. Right mouse button click
B. ‘B’ button press
C. ‘X’ button press
D. Red button press
E. Spacebar</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td></tr>
<tr><td>Would your enemy most likely be controlled by the computer or another person?
A. Computer
B. Person
C. Either Computer or Person</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td></tr>
<tr><td>The missile on the left side of the screen is about to hit which character (indicate the letter associated with the character)?</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td><td>○</td></tr>
</table>

13.

	A.	B.	C.	D.	E.	F.	G.	H.
Which controller from question 8 above would you most likely use with this game? A.  B.  C.  D.  E.  F.  G.  H. 	○	○	○	○	○	○	○	○
Would your enemy most likely be controlled by the computer or another person? A. Computer B. Person C. Either Computer or Person	○	○	○	○	○	○	○	○
How would you throw a pass to receiver Holt? A. Right mouse button click B. ‘B’ button press C. ‘O’ button press D. Red button press E. Spacebar	○	○	○	○	○	○	○	○

	A.	B.	C.	D.	E.	F.	G.	H.
Which controller from question 8 above would you most likely use with this game? A.  B.  C.  D.  E.  F.  G.  H. 	○	○	○	○	○	○	○	○
If you were controlling the character closest to you, what controller actions would you perform to defeat the enemies? A. Right mouse button click B. ‘B’ button press C. ‘X’ button press D. Red button press E. Spacebar	○	○	○	○	○	○	○	○
Would your enemy most likely be controlled by the computer or another person? A. Computer B. Person C. Either Computer or Person	○	○	○	○	○	○	○	○

Appendix C: Graphical User Interface Questionnaire

The table below presents the questions and response scales from the Graphical User Interface Questionnaire used during CMEX I. The table provides the question stem and end anchors of the response scale. Material added to the questions prior to the CMEX II administration are shown in parentheses and italics. Material deleted from the questionnaire used in CMEX I, for use in CMEX II, are underlined. The questionnaires were implemented for CMEX I using stand alone survey software, and converted to an internet format for CMEX II (with minor editing). The items used in the Fidelity, Avatar, Training, and Control Operations scales have the name in parentheses following the question stem.

Question	Response Scale
1. Please select the category that best describes your use of the system today.	(1) One of the Trainees (2) Acted as a Role Player (3) Exercise Trainer/Controller
2. Was the overall User Interface easy to understand and use? (Control Operations)	(1) Very Difficult (7) Extremely Easy
3. Does the User Interface seem(*s*) like a good design for this kind of simulation. (Control Operations)	(1) Strongly Agree (5) Strongly Disagree
4. Were the function keys easy to remember? (Control Operations)	(1) Very Difficult (7) Extremely Easy
5. Were the function keys easy to use? (Control Operations)	(1) Very Difficult (7) Extremely Easy
6. Was the movement control system easy to learn to use? (Control Operations)	(1) Very Difficult (7) Extremely Easy
7. Overall, which control type did you prefer?	(1) Letter Keys (2) Arrow Keys
8. Did the mouse control hinder or ease the movement and view control? (Control Operations)	(1) Very Difficult (7) Extremely Easy
9. Was it easy to move around in the environment AFTER you learned to use the controls? (Control Operations, Fidelity)	(1) Very Difficult (7) Extremely Easy
10. How realistic were the buildings/facilities? (Fidelity)	(1) Totally Artificial (7) Totally Real

11. Is experiencing collisions in the simulation important in moving an avatar around? (Fidelity)	(1) Very Important (5) Interfering
12. Please rate the building entry capabilities the system used in the simulation. (<u>for example, is there a problem in</u> (*i.e.*) not having functioning doors?) (Fidelity)	(1) Totally Artificial (7) Totally Real
13. How easy was it to recognize the avatars throughout the simulation?	(1) Difficult (5) Quite Easy
a. by physical features (Avatar, Fidelity)	
b. by voice (Avatar, Fidelity)	
c. at a distance (Avatar, Fidelity)	
14. Was it easy to detect collisions during movement (for example, hitting doorways during entry)? (Control Operations, Fidelity)	(1) Very Difficult (7) Extremely Easy
15. How realistic were the avatar's capabilities in these areas?	(1) Artificial (5) Totally Real
a. Movement (Avatar, Fidelity)	
b. Communication (Avatar, Fidelity)	
c. Making gestures (Avatar, Fidelity)	
d. Performing a visual inspection (Avatar, Fidelity)	
e. Performing a physical inspection (Avatar, Fidelity)	
16. Please provide a short description of any controls that did not work as you expected	Text Response
17. Please rate your agreement with the following statements:	(1) Strongly Agree (5) Strongly Disagree
a. Time to teleport was irritating. (Control Operations)	
b. Using the mouse to search objects is good enough for training. (Control Operations, Training, Fidelity)	
c. Using menus to enter vehicles or handle objects works fine for training. (Control Operations, Fidelity)	
d. Teleporting made it less real. (Fidelity)	
18. Individual avatars in the environment: (Avatar)	(1) were not very easily identified (5) were very easily identified
19. The appearance of the avatars in the environment: (Training, Avatar)	(1) will not support training (5) will enhance training effects
20. Can you remember any times when the system didn't keep up with what you were doing? How many times (0 - 10)?	Text Response
21. Was there any noticeable latency in the simulation? (Fidelity)	(1) System was very fast (5) System was too slow

22. Can you describe the worst interaction you had in the system? What were you doing?	Text Response
23. Which worked better, the "hands-free" or the "push-to-talk" voice control?	(1) Hands-free (2) Push-to-talk
24. What voice or communication capability needs to be improved or added to this system for general Army training?	Text Response
25. How good is the environment realism? (Fidelity)	(1) Very Poor (7) Extremely Good
26. How well was rank and authority reflected in the simulation?	(1) Totally Inadequate (5) Totally Adequate
a. Were indications of rank clearly available? (Fidelity)	
b. Were there indications of civilian status? (Fidelity)	
c. Was it possible to exercise authority to accomplish goals? (Training, Fidelity)	
27. What is the most important aspect of the visual displays to improve?	Text Response

Appendix D: Exercise Questionnaire

The table below presents the questions and response scales from the Exercise Questionnaire used during CMEX I. Changes or additions to the questions made prior to the CMEX II administration are shown in parentheses and italics. Material deleted from the questionnaire used in CMEX I, before use in CMEX II, are underlined. In combination, words or phrases replaced in CMEX II are indicated by underlined words followed by italicized words surrounded by parentheses. The questionnaires were implemented for CMEX I using stand alone survey software, and converted to an internet format for CMEX II. The items used in the Fidelity and Training Effectiveness scales have the name in parentheses following the question stem.

Question	Response Scale
1. Please select the category that best describes your use of the system today.	(1) One of the Trainees (2) Acted as a Role Player (3) Exercise Trainer/Controller
2. Please rate the avatar's capabilities based on your experience with the system?	(1) Totally inadequate (5) Totally Adequate for all actions
a. Movement (Fidelity)	
b. Communication (Fidelity)	
c, Gestures (Fidelity)	
d. Visual Inspection (Fidelity)	
e. Physical Inspection (Fidelity)	
3(*2*). How much did the animated gestures contribute to this exercise?	(1) Limited capabilities hindered activities (7) Capabilities were needed for (*supported*) many key activities
4(*3*). How does the simulation compare to field training exercises in:	(1) Much better (5) Much worse
a. (*The*) Diversity of tasks (Training)	
b. (*the*) Ability to record events for review & analysis (Training)	
c. (*the*) Time required for exercise (Training)	
d. (*the*) Ease of change in exercise (Training)	
5(*4*). Were there any important gestures that were not implemented, or was there a gesture that needed improvement (*for this exercise*)?	Text response
6(*5*). During the exercise(s), were there any important sounds:	yes/no
a. Missing? (Fidelity)	*(1) None* *(5) All*
b. Missing when expected? (Fidelity)	
c. Incorrect in characteristics? (Fidelity)	
d. Unexpected when they occurred? (Fidelity)	

(*6. During the exercise did any important sounds seem wrong?*) (Fidelity)	*(1) None* *(5) All*
7. Did the simulation exercise(s) require more or less preparation in the following areas?	(1) A lot more (5) A lot less
a. <u>than</u> a normal STX (Training)	
b. <u>than</u> a "walk through" preparation (Training)	
c. <u>than</u> a computer course (Training)	
8. Was there any noticeable latency in the simulation (*that affected the exercise*)? (Fidelity)	(1) System was (*always*) <u>very</u> fast (*enough for the exercise*) (5) <u>too slow</u> (*System was never fast enough for the exercise*)
(*9. Do you have any comments or suggestions about the system speed?*)	(*Text Response*)
<u>9</u>(*10*). Did the explosions and special effects seem real enough for training (*in these exercises*)? (Fidelity)	(1) <u>Too Hollywood</u> (*Too fake for any training*) (5) Good, will improve (*Soldier*) performance
<u>10</u>(*11*). Was the (*local*) voice system (*(not radios)*) adequate to support <u>the</u> (*this*) training exercise(s)? (Fidelity, Training)	(1) Inadequate (5) More than adequate (*for training*)
(*12. What single improvement in the local sounds presentation would most improve these kinds of training exercises?*)	(*Text Response*)
<u>11</u> (*13*). Please indicate your level of agreement with the following statements.	(1) Strongly agree (5) Strongly disagree
a. Once I got used to the simulation, I could easily focus on the necessary information for accomplishing my part of the mission.	(*(1) Strongly Disagree* *(5) Strongly Agree*)
b. It was easy to do most of the tasks called for in the exercise.	
c. The system performed as I expected.	
d. It was easy to correct any errors made during operation of the simulation.	
e. I did not make many errors in using the simulation.	
f. The difficulties in working with the simulation interfered with the exercise. (Training)	
g. Using the equipment interfered with conducting the exercise. (Training)	
<u>12</u> (*14*). What is the most important aspect of the visual display to improve?	Text Response

13 (*15*). How real or artificial were the following major aspects of the simulation?	(1) Artificial (5) Totally Real
a. Was the area of operations realistically scaled? (*Terrain area of operations*) (Fidelity)	
b. Was the transportation speed reasonable for training? (*Transportation speed (in a vehicle)?)* (Fidelity)	
c. Did you cross physical distances in realistic time? (*Physical movements (e.g. on foot)?*) (Fidelity)	
14 (*16*). Rate the areas below in terms of supporting teamwork to accomplish exercise goals. (*How well did each of the following areas support working as a team to accomplish the unit's mission in this exercise?*)	(1) Prevented (5) Enabled
a. Visual aspects. (Training)	
b. Gesture system capabilities. (Training)	
c. Communications capabilities (*aspects*). (Training)	
d. Movement system characteristics. (Training)	
15 (*17*). As a result of your experience in this exercise simulation, how do you think the average enlisted (*your*) Soldier capabilities would (*have*) change(*d*) from using this system? Please evaluate the potential change in the following areas.	(1) Much Worse (5) Much Better
a. Communication with Leaders. (Training)	
b. Communicate with other unit members. (Training)	
c. Gather the information necessary to support (*for unit*) decisions. (Training)	
d. Deal (*Negotiate*) with locals. (Training)	
e. Recognize hidden problems. (Training)	
f. Respond to Opposing Forces. (Training)	
g. Respond to IED situations. (Training)	
(*h. Understand and apply Rules of Engagement & Escalation of Force*) (Training)	

16 (*18*). Were the simulated radios appropriate (*adequate*) for the(*se*) scenarios? Please select the response that best reflects your opinion, or select "other" and enter a short (200 characters max.) comment in the blank. (Fidelity, Training)	(1) The radios didn't support the communications adequately (*in the exercise*). (5) The radios supported the needed communications extremely well, enabling focus on the training event. (6) Different equipment is primarily used for communicating information (e.g. BluFOR or FBCB2). (*How do the radios need to be changed for the exercise?*) (7) Other – text response.
17 (*19*). Please address the following issues by providing a rating using the five point scale.	(1) Inadequate (5)Great
a. The access to the binocular was:	
b. The binocular view when inspecting potential IED objects was: (Fidelity)	
c. The magnification by the binoculars was: (Fidelity)	
d. The binocular controls were:	
e. Using the binoculars from the vehicle was: (Fidelity)	
18. Was the simulation adequate for rehearsing or learning Cultural Understanding? (Training)	(1) The Simulation supported ALL aspects or activities of Cultural Understanding. (6) The exercise did not address cultural understanding.
19. Did the simulation adequately support needed gestures for the exercise? (Fidelity, Training)	(1) The Simulation supported ALL needed gestures. (5) The Simulation did not support any needed gestures.

20. What important physical capability was needed in this exercise that this simulation did not provide? (*What important physical capability in addition to the movements and gestures does this simulation need for these exercises?*)	Text Response
21. Was the simulation adequate for rehearsing or learning Escalation of Force and Rules of Engagement? (Training)	(1) The Simulation supported ALL EOF/ROE aspects or activities. (5) The Simulation did not support any EOF/ROE aspects or activities.
22. What one important capability is needed in order for this simulation to better support learning or rehearsal of EOF/ROE?	Text Response
23. Was the simulation adequate for rehearsing or learning Tactical Questioning during operations? (Training)	(1) The Simulation supported ALL aspects or activities. (5) The Simulation did not support any aspects or activities.
24. What one important capability is needed in order for this simulation to support rehearsal of Tactical Questioning?	Text Response
25 (*23*). Select the following statements that indicate your opinion about the vehicles used in the scenarios. Use the "Other" button to enter up to a 200 character comment. More than one selection is encouraged. (*Select all that apply*.)	Other – Text Response
a. The vehicles were inadequate for rehearsing these missions.	
b. The vehicles were adequate for conducting these missions.	
c. The vehicles needed to carry more people and equipment for these missions.	
d. A mix of vehicles would improve these missions.	
e. The missions really needed vehicles that allowed each Soldier to look out, use binoculars, and fire weapons.	

26. As a result of your experience in this exercise simulation, please evaluate the potential change in capability and understanding for the average Soldier from conducting similar exercises in the following areas.	(1) Much Worse (5) Much Better
a. Understanding Escalation of Force & Rules of Engagement. (Training)	
b. Capability in Tactical Questioning. (Training)	
c. Capability to negotiate with locals. (Training)	
27 (*24*). Please rate the following statements:	(1) Strongly Agree (7) Strongly Disagree (*(5) Strongly Disagree*)
a. As a team we currently like each other.	
b. My team members and I expect to like each other in the future.	
c. As a team we believe that it is important that the team members get along.	
d. As a team we feel that we are very similar.	
e. My team members and I feel that it is very important to socialize during the session.	
28. Please rate the following statements:	(1) Strongly Agree (7) Strongly Disagree
a (*f*). My team members and I were engaged in the task.	
b (*g*). As a team, we enjoyed the task.	
c (*h*). My team members and I agree that it is important to do well on the task.	
d (*i*). As a team, we felt that the task was meaningful.	
e (*j*). My team members and I expect that there will be benefits from our team's performance.	

25. Please indicate your level of agreement with the following statements:	*(1) Strongly Disagree* *(5) Strongly Agree*
a. The simulation system adequately supported the tactical movement of the unit.	
b. Members of the unit were able to conduct Intelligence, Surveillance, and Reconnaissance (ISR) tasks during the tactical movement.	
c. We were able to rehearse and improve our urban tactical movement patterns using this system.	
d. Our ability to identify and use overwatch positions was not well exercised in this simulation.	
e. The system supported analysis and reporting of possible enemy direct fire situations.	
f. The mounted resources could easily coordinate with the dismounted elements during movements and reactions.	
g. The simulation supported learning and improvement in understanding leader's intent and accomplishing individual tasks within the exercise.	
h. The system did not support use of terrain features in establishing fighting positions.	
i. The simulated terrain presented a challenge in setting up a checkpoint/roadblock that met all standard requirements.	
26. The most difficult aspect of tactical movement using this system was:	*Text Response*
27. How difficult or easy was establishing security in these exercises?	*(1) Very Difficult* *(7) Extremely easy*
28. How difficult or easy was it to employ vehicles during defensive operations in this simulation?	*(1) Very Difficult* *(7) Extremely easy*
29. How difficult or easy was establishing a hasty roadblock in these exercises?	*(1) Very Difficult* *(7) Extremely easy*

Appendix E: Biographical Questionnaire

Please enter todays date (YYMMDD): __________

PLEASE NOTE: The information gathered from these questions will not be attributed to any individual. No personal information will be released. The information will only be used in aggregate form during analyses, in order to relate responses to previous administrations, other questionnaires, and publicly available demographic data.

The first questions are about you and your job, followed by questions about your computer experience.

Please enter a unique number that you can remember easily. This number will only be used to track your questionnaire responses. (For example, the last 4 numbers of your phone are probably unique and easily remembered. This blank has a 10 character limit.) ____________

Please enter your age: _____

You are: ○ Male ○ Female

How many years have you been on active duty? (If not applicable, please enter a zero.) _____

Please enter your rank or grade. (There is a 50 character limit.)

____

Please enter the title or description of your Military Occupational Speciality (e.g. Military Police or Infantry). (50 character limit.)

____

Please identify your unit as completely as possible. (50 character limit)

____

Please describe your most recent deployment (location, dates/length of time, duties). (There is a 60 character per line, 300 character limit.)

____
____
____
____
____

Are you a trainer, or is training a major part of your job? (If not, please enter zeros for the nest two questions.) ○ No ○ Yes

How long have you been a trainer (in months and years)? ______________________

How many hours during the average week do you spend training others? (Please include preparation and execution time.) _____

How much expertise or experience in training others do you feel you have? (Please select the most appropriate level.)

- ○ Very Little Experience or Expertise
- ○ Some Experience
- ○ Below Average
- ○ Average Experience or Expertise
- ○ Above Average
- ○ Highly Experienced
- ○ Very Experienced or High Expertise

Do you supervise staff who spend any part of their time training others?

- ○ Yes
- ○ No

When did you start using computers?

- ○ younger than 6 years old
- ○ 6 to 11 years old
- ○ 12 to 14 years old
- ○ 15 to 17 years old
- ○ 18 to 20 years old
- ○ 21 to 23 years old
- ○ 24 to 29 years old
- ○ 30 to 39 years old
- ○ 40 to 49 years old
- ○ older than 50 years old

Please enter the average or typical number of hours per week that you use a computer. If you do not use a computer at all (on average), please enter a zero (0). Please use whole numbers in your estimate. (Maximum number allowed: 50) ____

Where do you currently use a computer?
Please select all that apply.

- ☐ Home, Barracks, or BOQ
- ☐ Unit Work Site
- ☐ Library, Learning Center, or Training Facility
- ☐ Other Site

Do you own a personal computer?

- O Yes
- O No

How often do you play computer games?

- O Daily
- O Weekly
- O Monthly
- O Less than once a month
- O Never

How often do you use icon-based programs or software?

- O Daily
- O Weekly
- O Monthly
- O Less than once a month
- O Never

How often do you play video games (run on a console, not a PC)?

- O Daily
- O Weekly
- O Monthly
- O Less than once a month
- O Never

How often do you use programs or software with pull-down menus?

- O Daily
- O Weekly
- O Monthly
- O Less than once a month
- O Never

How often do you use graphics or drawing features in software packages?

- O Daily
- O Weekly
- O Monthly
- O Less than once a month
- O Never

How often do you use email (at home or work)?

- O Daily
- O Weekly
- O Monthly
- O Less than once a month
- O Never

How often do you use the internet (not including email or gaming)?

- ○ Daily
- ○ Weekly
- ○ Monthly
- ○ Less than once a month
- ○ Never

How much do you enjoy playing video games (home or arcade)?

- ○ Not very much
- ○ Somewhat
- ○ Average enjoyment
- ○ A lot of fun
- ○ Most Fun in Life

Please rate your skill at playing video games.

- ○ Bad
- ○ Poor
- ○ Average
- ○ Better than Average
- ○ Good

Please enter the number of hours per week that you play video games. Please enter whole digits, e.g. 8 for eight hours. ____

How many times in the last year have you experienced a virtual reality game or entertainment?

○ Never	○ Six Times
○ Once	○ Seven
○ Twice	○ Eight
○ Three Times	○ Nine
○ Four	○ Ten
○ Five	○ More than 10 Times

What is your typing ability?

- ○ Hunt and peck slowly.
- ○ Hunt and peck quickly.
- ○ Type slowly while not looking at the keyboard.
- ○ Type quickly while not looking at the keyboard.

What is your level of computer expertise?

- ○ Novice.
- ○ Good with one type of software package.
- ○ Good with several different software programs.
- ○ Can program and use several software packages.
- ○ Expert.

Please indicate <u>all</u> of the types of software that you can use easily.

- ☐ Word Processing
- ☐ Spreadsheets
- ☐ Database
- ☐ Slides (like powerpoint)
- ☐ Scheduling / Calendars / Address list
- ☐ Audio Media (like iPod)
- ☐ Picture Media (like Photoshop)
- ☐ Movie Media (like Nero)
- ☐ Internet Browser
- ☐ Other (30 character limit): ______________________

In terms of writing programs or scripts, please check all of the languages that you have used or can easily use.

- ☐ None
- ☐ Visual Basic
- ☐ HTML
- ☐ XML
- ☐ Java
- ☐ C++
- ☐ Other: ______________________

How many hours have you spent training on equipment simulators (e.g. Firearms Training System, SIMNET, Convoy Training System, CCTT, etc.) in the last year? Please count only the hours spent using the simulator or simulation, not the associated time required for preparation, planning, or classroom work.

- ○ None
- ○ Less than 20 hours
- ○ 21 to 40 hours
- ○ 41 to 80 hours
- ○ 81 to 120 hours
- ○ More than 120 hours

Which of the following serious and/or military games have you used personally and/or to train with?

- ☐ America's Army
- ☐ DARWARS Ambush!
- ☐ Every Solider a Sensor Simulation (ES3)
- ☐ Blazing Angels 2: Secret Missions of WWII
- ☐ Halo 3
- ☐ World of Warcraft
- ☐ Full Spectrum Warrior
- ☐ Call of Duty
- ☐ Medal of Honor

Which of the following serious and/or military games have you used personally and/or to train with?

- ☐ Counter Strike
- ☐ Other: ____________________
- ☐ Other: ____________________
- ☐ Other: ____________________

Please describe your most recent simulation-based training. Please include the name or description of the system, its role in the training, and the training goal. (Do not include classroom exercises, games noted above, or simulated weapons used during a field exercise.)

____

____

____

Are there any simulators or simulations which you have used to conduct training for others?
Please identify the system or describe the training, and indicate the number of training sessions you have conducted.

____

____

____

Are there critical Soldier tasks on which you and your Soldiers currently do not receive enough training?
Please identify or describe the most important task.

____

____

____

Do you think a simulator or simulation could train those critical tasks?

○ YES ○ NO

Appendix F: After Action Review Questionnaire

The table below presents the questions and response scales from the After Action Review Questionnaire used during CMEX I. Changes to the questions made prior to the CMEX II administration are shown in parentheses and italics. Material deleted from the questionnaire used in CMEX I, before use in CMEX II, are underlined. The questionnaires were implemented for CMEX I using stand alone survey software, and converted to an internet format for CMEX II. The items used in the Interface and Training scales have the name in parentheses following the question stem.

Question	Responses
1. Please select the category that best describes your use of the system today.	(1) One of the Trainees (2) Acted as a Role Player (3) Exercise Trainer/Controller
2. Was the overall AAR Interface easy to understand? (Interface)	(1) Very Difficult (7) Extremely Easy
3. The User Interface for this AAR seems like a good design. (Interface)	(1) Strongly agree (5) Strongly disagree
4. Does it seem easy to move the point of view around in the environment during the AAR? (Interface)	(1) Very Difficult (7) Extremely easy
5. Was it easy to determine who was doing what during the AAR (Interface)	(1) Very Difficult (7) Extremely easy
6. Does it seem easy to move from event to event during the AAR? (Interface)	(1) Very Difficult (7) Extremely easy
7. Were the avatar's capabilities realistic enough for AAR use?	(1) Artificial (5) Totally Real
a. Movement (Interface)	
b. Communication (Interface)	
c. Gesture (Interface)	
d. Visual Inspection (Interface)	
e. Physical Inspection (Interface)	
8. How do the AAR capabilities compare to a field training exercise AAR in the following areas?	(1) Much Better (5) Much Worse
a. Presentation of tasks (Training)	
b. Ability to display Events (Training)	
c. Time required to conduct exercise AAR (Training)	
d. Ease of Preparation for AAR (Training)	
9. Were there any important AAR functions that were not implemented, or was there a capability that needed improvement?	Text response

10. During the AAR were there any important sounds:	No/Yes
a. Entirely missing? (Interface)	
b. Missing when expected? (Interface)	
c. Incorrect in characteristics? (Interface)	
d. Unexpected when they occurred? (Interface)	
11. Did the AAR require more or less preparation than a:	(1) A Lot More (5) A Lot Less
a. Normal STX? (Training)	
b. Map/terrains walk? (Training)	
c. Op ord (Training)	
12. Overall the AAR system seems to be easy to learn. (Interface)	(1) Strongly agree (5) Strongly disagree
13. Was the voice system adequate to support the AAR? (Interface)	(1) Inadequate (5) More than adequate
14. In general, could this AAR support Army training as it works right now? (Training)	(1) Incapable of Training (7) Could support all Tasks
15. What is the most important feature or capability needed by this system to better support a wide range of Army training and rehearsal?	Text Response
16. The AAR system made it easy to review and determine what happened in the simulation during the exercise. (Interface, Training)	(1) Strongly agree (5) Strongly disagree
17. The AAR system made it easier to determine which areas to focus upon during future exercises. (Training)	(1) Strongly agree (5) Strongly disagree
18. What types of training or rehearsal tasks do you think this simulation system (not just the exercises you have experienced) is BEST suited to support?	Text Response
19. What types of training or rehearsal tasks do you think this simulation system is LEAST suited to support?	Text Response